

# THIS WEEK

## EDITORIALS

**WORLD VIEW** Tie grants to reproducible results to raise standards p.7

**SCHIZOPHRENIA** Faulty brain wiring could hurt learning and reasoning p.8



**EVOLUTION** Why we see spiny seashells on the sea shore p.9

## A record made to be broken

*Japan's major science-funding agency has a clean record when it comes to research fraud. Now is the time for it to step up and resolve a long-running case of alleged scientific misconduct.*

Japan's major funder of big science, the Japan Science and Technology Agency (JST), is either very good at picking honest scientists, or does not look very hard for evidence to the contrary. The total number of fraud cases confirmed by the JST since its beginnings in 1957 — an organization that delivered ¥87.8 billion (US\$942 million) in research funding last year — is zero.

The US National Institutes of Health reports a dozen or so such cases every year. Elsewhere in Japan, there has been a spate of high-profile cases of fraud in recent years — including that of Yoshitaka Fujii, an anaesthesiologist who currently has the highest number of retractions by a single author. The JST's perfect record would seem to be against the odds.

A long-running case typifies the problems. Akihisa Inoue, retired president of Tohoku University in Sendai, has been hounded by accusations of scientific misconduct for years.

There is no evidence that Inoue has committed any misconduct, and he denies manipulating any data. He is one of the world's leading experts in metallic glasses — materials that are more elastic and more resistant to corrosion than metals — and has published more than 2,500 papers on the subject. He has previously told *Nature* that other researchers may simply lack the skills and experience to reproduce some of his lab's results (see *Nature* **470**, 446–447; 2011).

There is a clear need for the JST to resolve the situation. Questions over Inoue's work were first raised by an anonymous whistle-blower nearly six years ago, and the case touches on research going back two decades (see *Nature* **483**, 259; 2012). There have been many complicated twists. Tao Zhang, a co-author with Inoue on several papers,

says that the documents and metallic glass samples that could have settled the matter were lost when a container carrying his belongings back to his home in China fell into the sea. A committee set up by Tohoku University in 2007 to determine whether to investigate the misconduct allegations was dominated by university administrators who had previously been promoted to their posts by Inoue. The committee's conclusion — that there was no need to investigate — led to charges of conflict of interest by an economist and other colleagues at the university, who raised questions with journal editors. A series of corrections have been issued about the relevant papers.

Last year, a separate committee convened by the JST — which gave Inoue ¥1.8 billion for a five-year project starting in 1997, during which several of the problematic papers were produced — reported that although they couldn't find evidence of misconduct, a closer look was justified. So last spring, the JST formally requested that the university investigate. Twelve months on, neither the university nor the JST would confirm whether an investigation was under way, who would be on the investigation committee or when it would start and end.

There is no deadline by which any investigation must report back, and if Tohoku University fails to respond, the JST can take no action other than to ask again. The university, which might have to return some or all of the money if it discovers misconduct, again faces a conflict of interest in addition to the embarrassment of an investigation. Doing nothing apparently has no consequences. And it is one way, now and in the future, to ensure that the JST will keep its perfect record. ■

## Against the law

*Behaviours proposed for black holes conflict with the laws of physics.*

Never let it be said that *Nature* does not address the science questions that are on the lips of researchers. On page 20, a News Feature poses the conundrum: will an astronaut who falls into a black hole be crushed or burned to a crisp? The answer, according to the disappointing end to the Disney film *The Black Hole* (Gary Nelson; 1979) is neither: the astronaut will go on a psychedelic trip through a (possibly metaphorical) hell and heaven. The answer according to physicists in 2013 is even more of a let-down: they don't know.

The European particle-physics laboratory CERN, located near Geneva in Switzerland, has had a mixed experience with black holes: various challenges to its Large Hadron Collider over the years have

focused on the (very small) chance that it could create (very small) black holes that would destroy the world. So the CERN press office must have looked on nervously as physicists gathered there last month to address the mysterious fate of our unfortunate astronaut. There is more at stake than the grisly demise of a single free-floating space traveller. The physics is complicated but the take-home message is this: if the astronaut fries, then Einstein's framework of general relativity goes up in smoke with it; if the astronaut is crushed and torn by the black hole's internal variation in gravity, then quantum mechanics is wrong.

So which is it, relativity or quantum theory; heaven or hell? Debate continues. One compromise has the astronaut hang around outside the black hole to hoover up some quantum information as it leaks, use it to do some maths and then jump in to see if either theory is right. So far so good (for the laws of physics if not the astronaut, who perishes either way) except that the maths is so difficult that, by the time scientists have an answer, the black hole will have evaporated beneath them. Now that would make for a better ending, at least for the astronaut. Disney, which is said to be remaking its film, should take note. ■

JOCELYN FILLEY



## If a job is worth doing, it is worth doing twice

Researchers and funding agencies need to put a premium on ensuring that results are reproducible, argues **Jonathan F. Russell**.

In February, Facebook founder Mark Zuckerberg visited my university to announce the Breakthrough Prize in Life Sciences. As someone who hopes for a career in biomedical research, I applaud these awards. Biomedical science, when practised correctly, is a well-spring of knowledge, innovation and human health. But is it practised correctly? And if not, how does it need to change?

In recent years, it has become clear that biomedical science is plagued by findings that cannot be reproduced. This wastes grant money and hinders development of new treatments and cures. The irreproducibility epidemic is exacerbated by a funding structure that rewards publications above all else, whether they are reproducible or not. Science as a system should place more importance on reproducibility. Not every paper needs to be medically relevant, but at the very least they should all be reproducible. Reproducibility separates science from mere anecdote.

Some journals already offer to publish replication studies, and there are nascent projects aimed at reproducing work in individual fields from disease to psychology. But these are stop-gap solutions. A more comprehensive answer is required.

As a start, funding agencies should tie grant funding to replication. A portion of their budget could be set aside to pay for independent replication studies. If a paper cannot be replicated, the authors should be required to amend or retract it. Funding agencies would then consider a principal investigator's history of reproducibility in grant reviews. This key reform would ensure that career success rests on the validity of findings rather than whether they are published in trendy journals.

Many scientists will object to this proposal, and for many reasons.

*It sounds complicated.* Scientists would apply for grants, conduct research and publish their findings much as they do now. The only difference is that replication studies would be more plentiful and more would be published. Each replication would be electronically linked to the original paper and boost its validity. Studies that could not be reproduced would be marked as such. Not all non-reproducible work is flawed, of course, but a red flag against a paper, accompanied by an amendment from the original authors noting discrepancies in methodology and analysis that might explain why it was not reproduced, would become a normal part of scientific discourse. Science should work through open dialogue, accessible to everyone.

*It would slow the dissemination of novel research.* Only modestly, and the time lost would be trivial compared with the time currently lost chasing false leads.

*Who would do the replication studies?* People

are already doing them. Under my proposal, more of these studies would be published. More people will be needed, and if granting agencies provided funds, more scientists would do replication studies.

*Replication studies are hard.* Sharing reagents and methods is not that difficult, and it is already mandated by funding agencies. Unfortunately, many do not comply. My proposal would encourage original authors to work with replicating groups and make the studies easier, because each replication would bolster the original authors' reputation.

*Publishing amendments would take a lot of time.* It would certainly take more time than is currently spent on it, but scientists already have similar conversations by e-mail, at conferences and in lab meetings. If publishing amendments were explicitly rewarded, people would take the time to do it.

*Science already self-corrects.* True, but the timescale is long and the associated waste is vast. Even outright fraud can take decades to come to light, and negative results rarely come to light at all. Under my proposal, negative results would be valued and valuable.

*But there is no money available.* This reform should save money by redirecting spending towards science that is reproducible. Society will receive a higher return on its investment in the form of treatments and cures.

*This proposal is too radical.* The same was once said about open-access publication. Yet that is quickly becoming the norm thanks to policy changes by funding agencies.

*The US National Institutes of Health and other granting agencies will never adopt this policy.* National governments, which fund granting agencies, could impose the reform and ensure that agencies adopt it.

*Acknowledging irreproducibility in science undermines public trust.* Trust is already being undermined. It is best that we reverse the tide and restore that trust by practising transparency.

*It would discourage high-risk, high-reward science, which is less likely to be reproducible.* There would be no requirement that a study be replicated before publication. If a risky paper ends up being reproducible, everybody wins. But if repeated attempts to reproduce it fail, it is not a valid finding — there is no reward to society. Why pretend otherwise?

In 1953, then US president Dwight Eisenhower said that every dollar spent on war "signifies, in the final sense, a theft from those who hunger and are not fed, those who are cold and not clothed". The same is true for every dollar spent on science that cannot be replicated. We can do better. ■

NOT EVERY PAPER  
NEEDS TO BE  
**MEDICALLY  
RELEVANT**  
BUT THEY SHOULD  
ALL BE  
**REPRODUCIBLE.**

NATURE.COM  
Discuss this article  
online at:  
[go.nature.com/yuyrdi](http://go.nature.com/yuyrdi)

**Jonathan F. Russell** recently defended his PhD and is completing medical school at the University of California, San Francisco.  
e-mail: [jonathan.russell@ucsf.edu](mailto:jonathan.russell@ucsf.edu)

# RESEARCH HIGHLIGHTS

Selections from the scientific literature

## NEUROSCIENCE

### Faulty link in schizophrenia

The impaired learning and reasoning skills seen in people with schizophrenia may be caused by defective connectivity between two brain regions.

Joshua Gordon and Christoph Kellendonk at Columbia University in New York and their colleagues used mice to mimic the abnormal brain activation seen in people with schizophrenia when they perform cognitive tasks.

The researchers genetically engineered mice so that a synthetic compound would reversibly inhibit neurons in the mediodorsal thalamus. They found that even a slight reduction in activity of this brain region altered connectivity with the prefrontal cortex and caused poor cognition.

This implicates the altered brain patterns in humans with the disease as a contributor to rather than a consequence of cognitive deficits, the authors say.

*Neuron* 77, 1151–1162 (2013)

## EMULSION CHEMISTRY

### Russian-doll-style droplets

Complex emulsions, such as spheres of oil and water nested inside each other, can be assembled in one step using a simple device, report Chang-Soo Lee at Chungnam National



CHANG-SOO LEE



KIM TAYLOR/NATURE PICTURE LIBRARY

## ANIMAL COMMUNICATION

### Bees of bigger hives forage better

Larger honeybee colonies benefit from their greater capacity to gather information.

Matina Donaldson-Matasci at the University of Arizona in Tucson and her team prevented honeybees (*Apis mellifera*) from communicating through their waggle dance (pictured), and monitored what happened in different-sized colonies. Bees in large colonies in which normal communication occurred were the most efficient

food-finders, sending new foragers to known resources up to four hours earlier than smaller colonies or large colonies in which waggle dances did not convey information.

This work provides some of the first experimental evidence that communication is particularly beneficial to large groups of social insects, the authors say.

*Anim. Behav.* 85, 585–592 (2013)

## ANIMAL BEHAVIOUR

### Chemically camouflaged fish

In the dark and quiet of freshwater ponds, a fish may use a cloak of chemicals to prowl for prey.

William Resetarits at Texas Tech University in Lubbock and Christopher Binckley at Arcadia University in Glenside, Pennsylvania, set up mock ponds and recorded how several species of predatory fish affected whether aquatic beetles moved in, or frogs laid eggs. The authors placed the fish in screened chambers that concealed them visually, but not chemically, and measured colonization

of the ponds by more than a dozen prey species. The prey avoided ponds that contained fish such as bluegill (*Lepomis macrochirus*) and the bluespotted sunfish (*Enneacanthus gloriosus*), but ponds containing pirate perch (*Aphredoderus sayanus*) — a particularly rapacious beetle muncher — were almost as popular as ponds without fish, suggesting that this fish masks its scent.

Although chemical deception has been reported for specific predator-prey pairs, this perch may be the first example of generalized chemical camouflage, the authors say.

*Am. Nat.* http://dx.doi.org/10.1086/670016 (2013)

## QUANTUM INFORMATION

**Quality photons from nanocrystals**

Photons that are suitable for quantum communication can be emitted from solid materials.

A roadblock to building quantum information networks is the need for streams of identical photons that oscillate in sync with each other. Such streams are produced by lasers, but integrating these with electronics is difficult.

Now Mete Atatüre at the University of Cambridge, UK, and his colleagues report that they have generated such a stream from semiconducting nanocrystals, or quantum dots. These dots exhibit a phenomenon called resonance fluorescence, in which emitted photons resonate with one another, and so oscillate in sync.

*Nature Commun.* 4, 1600 (2013)

MARTIN HARVEY/GETTY

## MICROBIOLOGY

**Whole sequence, no culture**

The development of a culture-free method for sequencing the genomes of chlamydia-causing bacteria means that researchers can obtain genomes directly from clinical samples and skip the tricky, time-consuming process required to grow these bacteria.

Helena Seth-Smith and Nicholas Thomson at the Wellcome Trust Sanger Institute in Hinxton, UK, and their colleagues purified *Chlamydia trachomatis* (pictured) from genital swabs using antibody-encrusted magnetic beads that latch onto *C. trachomatis* cells — leaving

EYE OF SCIENCE/SPL

behind contaminants such as human cells and other microbes. Bacterial DNA was then copied, sequenced and assembled using bioinformatic software; this yielded five complete genome sequences from 18 samples tested.

The technique could also be used to sequence other difficult-to-culture pathogens, the authors say.

*Genome Res.* <http://dx.doi.org/10.1101/gr.150037.112> (2013)

## NEUROSCIENCE

**Fickle wiring in ageing brains**

The decline of cognitive ability seen in some elderly people may not result from fewer brain cells, but from a rise in capricious connections.

Vincenzo De Paola at Imperial College London, Sen Song of Tsinghua University in Beijing and their colleagues observed the cell connections that shoot signals through the brain by fitting glass windows into the craniums of old and young mice. Although the researchers expected older mice to have fewer firing points, they found more. But these connections were erratic. In older mice, the connections had high turnover rates, and wavering firing strength. The older mice also performed less well on a memory test.

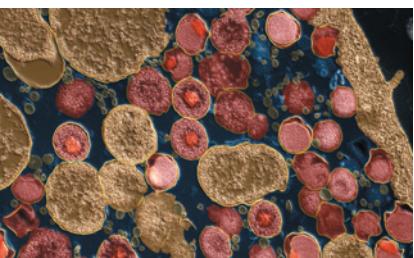
The researchers suggest that mental decline seen in ageing mice is due to disorderly wiring. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1218731110> (2013)

## PARTICLE PHYSICS

**Symmetry affirmed**

An unprecedented measurement of an antiproton suggests that matter and antimatter truly mirror each other.

Gerald Gabrielse at Harvard



## MATERIALS SCIENCE

**Making films on the edge**

★ HIGHLY READ  
on www.acs.org  
in February

A change in the orientation of molecular sheets in layered films may open up new uses for these materials.

Thin films are usually grown by stacking atoms horizontally, like sheets in a pad of paper. But Yi Cui and his colleagues at Stanford University in California have grown layers vertically, like dominoes balanced on their edges. The researchers started with strips of molybdenum and added sulphur or selenium atoms, which diffuse down through the molybdenum, carving nanometre-sized stripes.

The molybdenum atoms are sandwiched between sulphur or selenium atoms so that the top of the structure is made up of exposed edges. These sheet-edge structures are reactive, and so might make good catalysts, the authors say. As an example, they show that a reaction to produce hydrogen gas proceeds more rapidly with a higher concentration of sheet edges.

*Nano Lett.* 13, 1341–1347 (2013)



University in Cambridge, Massachusetts, and his group began with a particle trap designed to measure the proton's magnetic moment — the amount of force that its magnetic spin can exert. They shipped the equipment to the world's largest particle-physics laboratory — CERN near Geneva, Switzerland — where they confined a single antiproton from the CERN beam and measured the strength of its magnetic moment. The moment of the antiproton differed from that of the proton by less than 5 parts per million, the most precise measurement ever achieved. This result is in agreement with established theory of symmetry between particles and their antimatter counterparts.

*Phys. Rev. Lett.* 110, 130801 (2013)

independently in different species. Now, their shape can be explained by the physics of shell formation.

Derek Moulton at the University of Oxford, UK, and his colleagues modelled the interactions between the hard edge of the shell and the mollusc mantle that secretes new shell material during growth. Variations in growth rates and the stiffness of new shell could account for the diversity of spine shapes found on the shells of different species in sea-snail families such as Muricidae (pictured).

A mechanical model could explain the similarities in shell spines seen across various species, the authors suggest.

*Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1220443110> (2013)

## APPLIED MATHS

**Mechanics behind seashell spines**

Spines on seashells, which offer protection from the crushing jaws of potential predators, are often cited as an example of similar features that have evolved

© NATURE.COM  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

## POLICY

### Desert treaty woes

Canada withdrew from the United Nations Convention to Combat Desertification (UNCCD) on 28 March, leaving 193 countries party to the legally binding treaty. Canada is due to pay around US\$310,000 to the convention in 2013 — just over 3% of the convention's budget. The nation will honour its 2013 financial commitments, say media reports. The move comes ahead of conferences to be held under the aegis of the convention this month; where the first ever cost-benefit analysis of desertification and sustainable land management will be presented.

### UK open access

Rules making publicly funded research in the United Kingdom free to read came into force on 1 April. This year, universities and research institutions are expected to openly publish at least 45% of any research funded with grants from the UK research councils. Results will be made available either immediately or after a delay of between 6 and 24 months, depending on the research subject. This policy is due for review in 2014. See [nature.com/scipublishing](http://nature.com/scipublishing) for *Nature*'s special on open access and the future of publishing.

### China spring-clean

Beijing is spending 100 billion renminbi (US\$16 billion) on a three-year effort to clean up the city's air, land and water. The plan, announced on 28 March, includes measures such as shutting down illegal construction sites, setting up water-recycling plants and establishing an emergency headquarters to deal with extreme air-quality problems. Shanghai launched its own air-pollution plan on



AURORA PHOTOS/ALAMY

## US waterways in bad shape

More than half of US rivers and streams are in a poor environmental condition, according to a survey published by the Environmental Protection Agency on 26 March. The data from 2008–09 — the most recent available — show that 28% of the nation's waterways have excessive levels of nitrogen, and 40% have high levels of phosphorus. This nutrient pollution

causes algal blooms (pictured; near the Iron Gate dam on the Klamath River in northern California), the breakdown of which saps water of the oxygen that aquatic life needs to survive. "This is the most dire in a string of water assessments over the last 20 years," says Earthjustice, an environmental law organization in San Francisco, California.

1 April, providing emergency measures for when air quality declines below a set target. These include halting building projects and swapping coal for natural-gas in power generation.

### Climate change

More than 80% of Americans believe that the planet is warming and think that coastal communities should be better prepared for rising seas and stronger storms, a survey published on 28 March has found. However, the poll, commissioned by scientists at Stanford University in California, also found that the US public does not want the government to pay for climate-change adaptation. Most of

the 1,174 people surveyed said that those living in at-risk areas should bear the costs of making their communities more resilient. See [go.nature.com/z23dot](http://go.nature.com/z23dot) for more.

## RESEARCH

### HeLa-cell debate

An ethical controversy erupted last week over whether to publish genetic data on old tissue samples that were originally established without their donor's consent. Researchers hurriedly removed from public databases their recently published genome of the world's most famous cell line, HeLa, which was established without consent from

Henrietta Lacks around 1951 (when consent was not legally required). Lacks's descendants, together with bioethicists and other scientists, said that it could reveal information about traits carried by living family members — although other genetic data about HeLa cells remain available online. See [go.nature.com/yimiwi](http://go.nature.com/yimiwi) for more.

## FACILITIES

### Deep-sea dive

The film-maker James Cameron is donating his deep-sea submersible to the Woods Hole Oceanographic Institution in Massachusetts, he announced on 26 March. Cameron used the craft, the

**DEEPSEA CHALLENGER**, to set a record last year for the deepest single-person dive when he travelled to the bottom of the Marianas Trench, reaching a depth of 10,900 metres. The *DEEPSEA CHALLENGER* is the only submersible capable of carrying a human deeper than 7,000 metres. See [go.nature.com/hhsks8](http://go.nature.com/hhsks8) for more.

## EVENTS

**Bird flu deaths**

Two men have died after being infected with a type of bird flu never before seen in humans, China reported on 31 March. The men, aged 27 and 87 and from Shanghai, contracted the H7N9 virus in February and died within 8 days. A 35-year-old woman from Anhui province, who contracted the virus in March, was in a critical condition as *Nature* went to press. See [go.nature.com/vrjulj](http://go.nature.com/vrjulj) for more.

**Locust plague**

The worst locust plague to hit Madagascar in decades prompted the United Nations on 26 March to call for US\$41 million in funding from donors to implement immediate emergency and longer-term control measures. Around half the country is already affected by the pests (**pictured**), which threaten large swaths of the island's rice



production and could thrust more than 10 million residents into hunger. Called-for control efforts include aerial spraying of some 1.5 million hectares and improved aerial and ground surveillance to monitor the threat.

## BUSINESS

**Patent blocked**

India's Supreme Court ruled on 1 April against an attempt by Swiss drug company Novartis to patent the anticancer drug Gleevec (imatinib mesylate). India's patent law limits the ability of drug firms to make minor modifications to drugs to extend their patent life, a tactic known as evergreening. The patent claim by Basel-based Novartis on a modified version of Gleevec "fails in both the tests of invention and patentability", India's high court said. The nation's support for cheap generic drugs has reduced prices, notably those

of anti-HIV medicines. But drug companies complain that it weakens intellectual-property rights and stifles innovation. See [go.nature.com/csqli](http://go.nature.com/csqli) for more.

**GM crop deal**

Two agricultural biotechnology giants agreed on 26 March to settle a lengthy legal battle over patent rights to 'next-generation' genetically modified (GM) soya beans. DuPont, based in Wilmington, Delaware, will pay Monsanto, based in St Louis, Missouri, at least US\$1.75 billion over the next decade for the right to offer two herbicide-tolerant lines of soya bean. Monsanto will gain access to some DuPont patents covering disease resistance and maize (corn) defoliation.

**Diabetes drug**

US regulators have given the green light to the first in a new class of drugs to treat diabetes. On 29 March, the Food and Drug Administration approved canagliflozin (Invokana), which was developed by Johnson & Johnson, headquartered in New Brunswick, New Jersey. Canagliflozin blocks sugar reabsorption in the kidneys by inhibiting the sodium/glucose cotransporter 2 (SGLT2) protein, increasing the amount of sugar patients excrete in their urine. Companies have

## COMING UP

**6–10 APRIL**

Discussions on mapping the epigenomes of cancer take place at the American Association for Cancer Research's annual meeting in Washington DC. [go.nature.com/wiozbm](http://go.nature.com/wiozbm)

**7–11 APRIL**

Plenary talks about the sustainability of the world's food system feature at the American Chemical Society's Spring Meeting in New Orleans, Louisiana. [go.nature.com/xgylsf](http://go.nature.com/xgylsf)

**8–19 APRIL**

The United Nations Forum on Forests meets in Istanbul, Turkey; it will discuss progress on a 2007 commitment to manage forests sustainably. [go.nature.com/potx5i](http://go.nature.com/potx5i)

been racing to bring an SGLT2 inhibitor to market; last year regulators shot down another such drug, dapagliflozin, citing cancer concerns.

## PEOPLE

**NICE head departs**

The head of the National Institute for Health and Care (formerly Clinical) Excellence (NICE) in London stepped down on 31 March. Michael Rawlins has led the body, which is tasked with assessing the cost-effectiveness of medicines, since 1999. The government confirmed last month that NICE will have a central role in a new pricing scheme for medicines under which, from 2014, drug companies will be paid what their products are deemed to be worth. For more see [go.nature.com/37lsu2](http://go.nature.com/37lsu2).

## NATURE.COM

For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

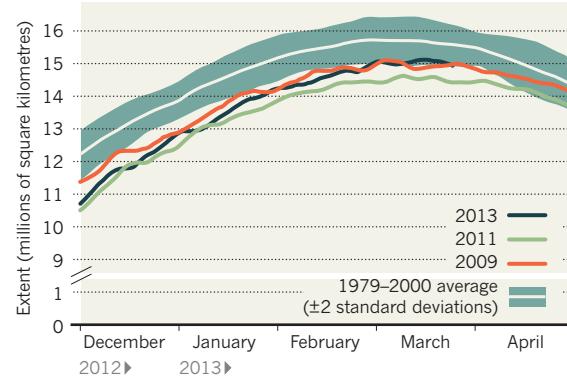
## TREND WATCH

SOURCE: US NATIONAL SNOW AND ICE DATA CENTER

A record low in the extent of sea ice in the Arctic last September has been followed by a record refreezing of uncovered ocean surface, resulting in a winter maximum on 15 March that is still the sixth-lowest recorded since satellite measurements began in 1979. The US National Snow and Ice Data Center in Boulder, Colorado, said that the data indicate "a more pronounced seasonal cycle" and "the increasing dominance of first-year ice in the Arctic".

**ARCTIC ICE MAXIMUM**

Sea-ice extent in the Arctic has reached its sixth-lowest maximum in the satellite record.



# NEWS IN FOCUS

**BIOMEDICINE** Tracking a slow, steady immune response to cancer **p.14**

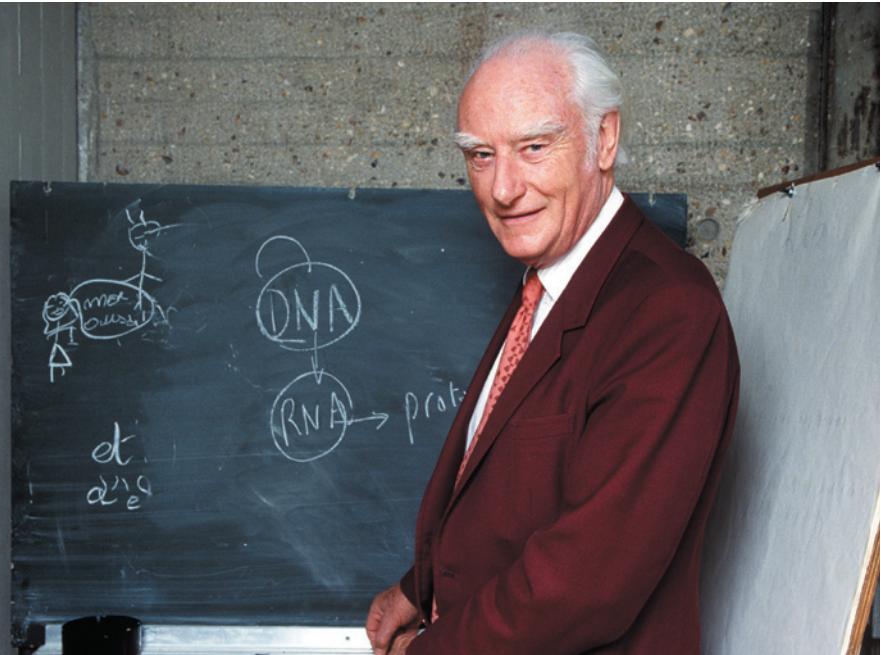
**GEOPHYSICS** Neutrino detectors monitor the fire in Earth's belly **p.17**

**FISHERIES** Study points to missing megatonnes in China's catch **p.18**



**BLACK HOLES** Astronaut and Einstein both get singed in debate **p.20**

DANIEL MORDZINSKI/AFP/Getty Images



Francis Crick's co-discovery of the structure of DNA laid the foundations of molecular biology.

SCIENCE HISTORY

## Crick's medal goes under the hammer

Pending auction raises eyebrows but few objections.

BY EWEN CALLAWAY

**B**efore he and James Watson made the defining discovery of the twentieth century, Francis Crick was an impoverished PhD student. The future Nobel laureate twice left his typewriter with a pawnbroker in exchange for a loan, and his family helped Crick and his wife Odile to make ends meet.

Now he is — posthumously — repaying that generosity. Crick's family stands to gain more than a million dollars after his Nobel medal and other memorabilia go to the auction block next week in New York. The sale has been met with speculation and curiosity, but also with some shrugs. Historians are less interested in

the glittery prize that Crick won six decades ago than in some faded bits of paper that are also up for bidding: a letter from Crick to his son describing the landmark 1953 discovery.

Public attention, however, has focused on Crick's Nobel medal since its 11 April sale was announced. The 23-carat-gold medallion was awarded for the 1962 Nobel Prize in Physiology or Medicine, which Crick shared with Watson and Maurice Wilkins for their work on determining the molecular structure of DNA and its role in heredity. Crick put the medal in a family trust in 1990.

Heritage Auctions, a

NATURE.COM

For more on Crick's memorabilia and the full letter to his son:  
[go.nature.com/fucsxk](http://go.nature.com/fucsxk)

firm based in Dallas, Texas, that specializes in coins and other historical memorabilia, estimates that the medal will fetch more than US\$500,000. The company is also selling an endorsed cheque in Swedish kronor awarded for the prize; a lab coat perplexingly adorned with a single — not double — golden helix; several of Crick's books; and his gardening journal and nautical logbook. (He was an avid gardener and used his Nobel money to buy a half-share in a 14-metre yacht.)

One day before this sale, auctioneer Christie's will sell the seven-page, handwritten letter that Crick sent to his 12-year-old son Michael in 1953 describing the structure of DNA. "Jim Watson and I have probably made a most important discovery," begins the letter, which includes a sketch of the double-helical structure and Crick's inference on how the molecule replicates. "You can understand that we are very excited."

Christie's estimates that the letter will fetch between \$1 million and \$2 million. "It's a home-run kind of letter," says Francis Wahlgren, international head of rare books and manuscripts at the auction house. In assessing the letter's worth, he cites its scientific importance — the letter is one of the first written descriptions of DNA's structure, preceding its scientific publication by weeks — as well as its personal sentiment. Half of the letter's sale money will be donated to the Salk Institute for Biological Studies in La Jolla, California, where Crick ended his career. And at least 20% of the money from the sale of the medal and the other items will benefit the Francis Crick Institute, a biomedical centre under construction in Central London.

Kindra Crick, a granddaughter of the laureate who is an artist in Portland, Oregon, says that the medal has been in storage for decades, and that her family hopes a buyer will make it available for public display. "It's challenging to know what to do with something like this," she says. "We pretty much figured that the types of people who would want to bid on this unique part of science history would have the resources to make it available for public display and not keep it locked up."

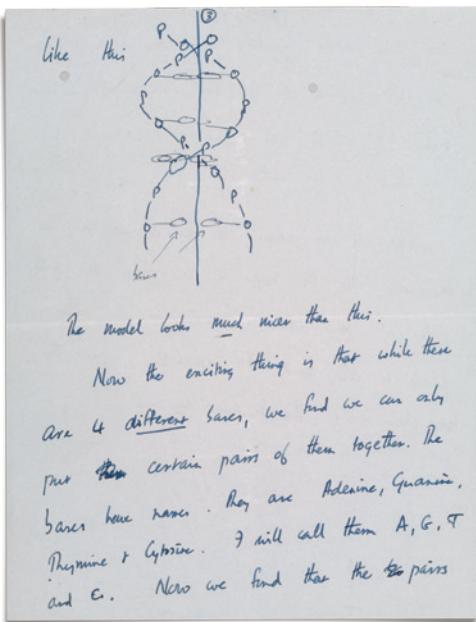
"That is not at all guaranteed," says Soraya de Chadarevian, a science historian at the University of California, Los Angeles. Still, she sees the medal as more of a collector's item than an important artefact from the dawn of molecular biology. The letter to Crick's son would

► be of much more interest to historians, she says, because it contains an early sketch of DNA. That, she adds, "would be much better placed in an archive".

In the early 2000s, Crick nearly sold his papers to a private buyer. Watson urged him to reconsider the sale, and in 2001, the Wellcome Trust, a London-based biomedical charity, purchased the papers for £1.6 million (US\$2.4 million). The Wellcome Library recently launched a website containing digitized versions of the collection.

Simon Chaplin, director of the Wellcome Library, says that the Crick memorabilia now at auction are a "low priority" for Wellcome. "I think it would be nice if the medal could be available to the public, but I feel it's more important to present the substance of somebody's research than to focus on individual items that may be intrinsically valuable," he says. He suggests that the medal and other items would be better suited to London's Science Museum: its collection includes a reconstruction of Watson and Crick's first model of the DNA double helix, complete with some pieces from the original model, which was destroyed.

Robert Bud, the museum's principal curator of medicine, says "it's our business whether



"A most important discovery," wrote Crick to his son.

we're going to bid for it", but that "if somebody were to give it to the Science Museum I would doubt we would reject it". Museums, adds Bud, can make scientific memorabilia more worthwhile by presenting objects within a larger story. "It makes the past real, and enables you

to believe that the future with which you're working has a real ancestry," he says.

It is uncommon — although not unheard of — for scientific collectables to wind up at auction, adds Bud. In 1997, his museum bought a preserved penicillin mould, prepared by Alexander Fleming, for around \$25,000 from Christie's. And in November 2012, a Danish auction house garnered 280,000 Danish kroner (US\$48,000) for the Nobel medal that Aage Niels Bohr won in 1975 for his theoretical work on the structure of the atomic nucleus. His father Niels Bohr, a pioneer of the field of quantum physics, and August Krogh, the 1920 medicine laureate, auctioned their medals in 1940 to raise money for a fund to aid the Finnish people during the Second World War. An anonymous buyer has since given the medals to the Danish Historical Museum in Fredriksborg.

Neuroscientist Charles Stevens, a former colleague of Crick's at the Salk Institute, isn't bothered by the family's decision to auction the medal. "The things that I would treasure from him are things he actually did," he says. Nor does he think that Crick would mind. "Francis didn't make the medal — he just won it. I'm sure he was glad he got it, but I doubt he would be very sentimental." ■

**BIOMEDICINE**

but early results are converging on one point: that patients' own immune responses to cancer are crucial in determining outcomes.

"When you sum it all up in many different cancers, early and late stage, it looks like patients who have a 'smouldering' initial immune response to cancer are more likely to benefit from immune therapies," says Thomas Gajewski, a cancer researcher at the University of Chicago in Illinois. Early work suggests that such patients may fare better, irrespective of the therapy used to treat them, he says. The evidence is fuelling a push to add an immune component to standard tumour pathology evaluations (see 'T cells can brighten prognosis').

***"It looks like patients who have a 'smouldering' initial response to cancer are more likely to benefit."***

Reliably gauging the initial response, however, is a challenge, cautions Mario Sznol, an oncologist at Yale University in New Haven, Connecticut. The immune response is a moving target, influenced not only by a patient's genetic make-up and environment, but also by proteins secreted by the tumour that can suppress immune cells. Immune markers present in the blood may not reflect what is happening in the tumour, and variation within a tumour means that biopsies may not paint a full picture either, if the needle misses

# Sizing up a slow assault on cancer

*Rise of immunotherapies spurs search for markers of response.*

BY HEIDI LEDFORD

Jedd Wolchok braced himself as he walked into the examination room to deliver bad news to his patient. Scans showed that the man's advanced melanoma had spread, and new tumours had sprouted, even though he had received an experimental therapy called ipilimumab (Yervoy) to rally his immune system against the disease. "In my mind I was rehearsing the standard speech," says Wolchok, an oncologist at the Memorial Sloan-Kettering Cancer Center in New York. "I'm sorry Mr So-and-so, but the treatment doesn't seem to be working." Instead, the patient stopped Wolchok at the door. "Now before you show me your pictures, let me tell you: I feel better," he said.

Wolchok took a gamble, and continued the therapy. More than six years later, his patient is thriving, and the drug, made by Bristol-Myers

Squibb of New York, has been approved by the US Food and Drug Administration. In that time, other experimental immunotherapies have shown promise, and clinicians have become increasingly familiar with delayed responses such as the one Wolchok's patient experienced — which reflect a gradual attack by the immune system, in contrast to the rapid onslaught of toxic chemotherapies. They still have no tools to quickly discern which patients are benefiting from treatment, but that may be about to change.

At the annual meeting of the American Association for Cancer Research in Washington DC this week, Wolchok and other researchers will report on their search for immunotherapy markers — ways to predict a patient's response to an immunotherapy or to show whether a given treatment is working. The work is hampered by the complexity of the immune system,

## CANCER STAGES

### T cells can brighten prognosis

For more than half a century, pathologists have assessed the severity of a cancer using the same basic framework: tumour size, the cancer's presence in lymph nodes and whether it has spread to other parts of the body. Called 'TNM staging', the framework ranges from stage 1, which has a relatively good prognosis, to stage 4, which can require immediate and aggressive therapy.

But researchers are now wondering whether it is time to add an 'i' category to the test — by factoring in whether immune cells are present in or around the tumour. In 2006, Jérôme Galon of the National Institute of Health and Medical Research in Paris and his colleagues found that the presence of T cells in or around colon cancers was associated with better patient outcomes<sup>2</sup>. They then applied the screen to samples from 599 patients and found that the immune response predicted outcomes more

accurately than the TNM staging system<sup>3</sup>.

The results have spawned an effort to test the 'immunoscore' approach at 23 sites in 17 countries. Researchers will mine stored specimens and associated hospital records for 6,000 colon-cancer patients. Results are expected in the autumn. Bernard Fox, who studies immunotherapies at Oregon Health & Science University in Portland, hopes that the test will also work for other cancers in which T cells can infiltrate tumours, such as breast and ovarian cancer. But he worries that biopsies for these other cancers are often not as large — in colon cancer, the entire tumour plus surrounding tissue is typically taken.

Yet if the colon-cancer results hold up at scale, it could prompt a radical change in thinking, he says. "In the current staging system, pathologists consider immune cells a contaminant," says Fox. "Isn't it incredible that they may be the key to survival?" H.L.

any patches of immune cells that have infiltrated the tumour.

Some are hoping that new methods will fill the gap. Adaptive Biotechnologies, based in Seattle, Washington, uses intensive DNA sequencing of blood and tissue samples to profile the changes in specialized receptors found in populations of immune cells called T cells. These cells make the receptors by shuffling variable regions of DNA in response to antigen exposure. Chief executive Chad Robbins says that the company hopes the assay will be sensitive enough that blood samples alone could be used to monitor patients' responses to immunotherapy. Demand for the technique has skyrocketed in the past year, he says, as more and more companies seek to test its potential to guide immunotherapy decisions.

Suzanne Topalian, an oncologist at Johns Hopkins University in Baltimore, Maryland, focuses instead on the cancer cells. A number of the world's biggest pharmaceutical companies are developing therapies that release a brake on the immune system by inhibiting a protein called programmed death 1 (PD-1) or its

partner, programmed death ligand 1 (PD-L1). A study by Topalian and her colleagues<sup>1</sup> suggests that these therapies work best if at least 5% of a patient's biopsied tumour cells express PD-L1. Some companies have already made PD-L1 expression a requirement for participation in their clinical trials — a move that both Topalian and Wolchok worry is premature. "I certainly would not want to exclude a patient from a trial for a PD-1-pathway blocking drug," says Wolchok. "Even if one tumour biopsy is negative, the patient may have other tumours that are positive."

While Topalian and others hunt for useful immunotherapy markers, academic labs are trying to understand what ignites the smouldering immune response to cancer in the first place. "Why do some patients have this response when others don't?" asks Sznoł. "Understanding this could lead us to new therapies." ■

- Topalian, S. L. et al. *N. Engl. J. Med.* **366**, 2443–2454 (2012).
- Galon, J. et al. *Science* **313**, 1960–1964 (2006).
- Mlecnik, B. et al. *J. Clin. Oncol.* **29**, 610–618 (2011).



**MORE  
ONLINE**

#### TOP STORY



Bioethics brouhaha over published HeLa genome [go.nature.com/ymiwgi](http://go.nature.com/ymiwgi)

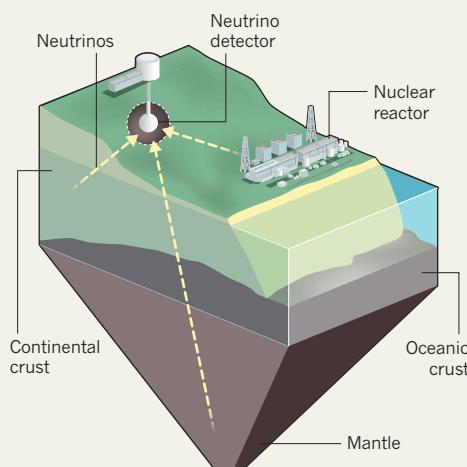
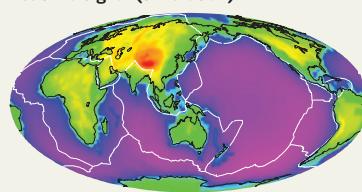
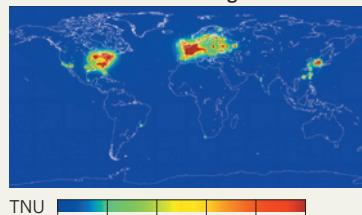
#### MORE NEWS

- More warming, more ice: an Antarctic paradox [go.nature.com/lo9axj](http://go.nature.com/lo9axj)
- Dinosaur colours in dispute [go.nature.com/zjmjrz](http://go.nature.com/zjmjrz)
- Wind, not water, may have built Mars rover's mountain target [go.nature.com/glgwiq](http://go.nature.com/glgwiq)

SOURCES: ONDŘEJ ŠRÁMEK, UNIV. MARYLAND; BARBARA RICCI, INFN

**UNDER THE SEA**

Researchers want to put a geoneutrino detector on the ocean floor to avoid confusing signals from nuclear reactors and Earth's crust.

**Combined crust and mantle neutrino signal (simulation)****Nuclear reactor neutrino signal**

TNU: Terrestrial neutrino units. 1 TNU corresponds to 1 geoneutrino event recorded over a year-long fully efficient exposure to  $10^{22}$  free protons.

2002 and November 2012 (ref. 3).

That's just enough for researchers to start drawing conclusions about the composition of Earth's mantle, says McDonough. Assuming that uranium and thorium are spread uniformly in the mantle, the KamLAND findings suggest that about 11 of the 47 TW come from the radioactive decay of those elements. A similar calculation for Borexino yields about 18 TW.

Ultimately, geoneutrino researchers would like multiple detectors spaced around Earth, so that they could perform a sort of tomography on the mantle. That could help scientists to discern between models that favour the uranium and thorium being spread throughout the mantle, versus those in which the elements are concentrated near the core–mantle boundary. Such a difference could help to determine where and how long heat will continue to flow to drive geological processes such as plate tectonics — and how long it will take Earth to cool.

One challenge is that emissions from uranium and thorium much nearer the surface in the continental crust can mask the geoneutrino signal coming from deeper in the planet (see 'Under the sea'). Next year, for example, the retrofitted Sudbury Neutrino Observatory (SNO) in Ontario, Canada, will start taking data with a 780-tonne detector that is sensitive to geoneutrinos. But SNO+, as the upgrade is called, sits smack in the middle of continental crust. Separating crustal from mantle geoneutrinos is crucial, says Steve Dye, a physicist at Hawaii Pacific University in Honolulu, as "the mantle is really what contributes to the rate of cooling of the planet".

Dye and others say that the best way to catch mantle geoneutrinos would be from the ocean floor, where the crust is thinner than on land. One scheme, dubbed Hanohano, would lower a 10,000-tonne detector from a barge, and has been on the drawing board for years. Construction alone would cost some US\$50 million to \$60 million, says John Learned, a neutrino physicist at the University of Hawaii at Manoa in Honolulu, and the technology is ambitious.

"We've never done anything like this before," he says. But interest in the project is growing, he adds, and supporters are trying to drum up funds to keep it moving.

Meanwhile, China is working on its Daya Bay II experiment, a 20,000-tonne detector on land that could be ready to hunt for geoneutrinos in 2019. Borexino has funds to run for at least another four years. And KamLAND plans to keep going for at least five more years, says team member Hiroko Watanabe of Tohoku University in Sendai, Japan. Even after Japan's nuclear reactors restart, the detector will still be able to find geoneutrinos — just not as easily. ■

**GEOSCIENCE**

# Detectors zero in on Earth's heat

*Geoneutrinos paint picture of deep-mantle processes.*

BY ALEXANDRA WITZE

A window on the deep Earth opened unexpectedly in 2011, when Japan's nuclear reactors were shut down after the Fukushima disaster. Before the closure, an underground particle detector called KamLAND based in Kamioka, Japan, was monitoring a torrent of neutrinos streaming from dozens of nearby nuclear reactors, seeking clues to the nature of these hard-to-catch subatomic particles. After those plants fell silent, KamLAND scientists could see more clearly a signal that had largely been obscured: a faint trickle of neutrinos produced inside the planet.

Neutrinos are generated in stars, reactors, and deep in Earth's crust and mantle by the radioactive decay of elements such as uranium and thorium. KamLAND reported the first tentative detections of these 'geoneutrinos' in 2005 (ref. 1). But last month at a conference in Takayama, Japan, KamLAND scientists reported seeing them in meaningful quantities — as did a team at the Borexino neutrino detector at the Gran Sasso National Laboratory near L'Aquila, Italy.

These detections are not just curiosities. Geoneutrinos offer the only way to measure one of Earth's internal heat sources. The total

heat flow, measured with sensors in deep mines and amounting to 47 terawatts (TW) of power, drives everything from plate tectonics to Earth's magnetic field. Some of it comes from the decay of radioactive elements, the rest is primordial heat left over from when Earth was formed by the violent collision of planetary building blocks.

But no one knows the proportions. Geologists assume that Earth contains the same amount of radioactive elements as certain primitive meteorites, but they aren't sure. "We're after trying to understand how Earth was built," says William McDonough, a geologist at the University of Maryland in College Park.

Enter KamLAND and Borexino, which spot geoneutrinos as a sideline to their other neutrino studies. Both experiments use liquid scintillator detectors, in which huge vats of fluid capture the occasional sparkle of light when a passing neutrino interacts with atomic nuclei in the liquid.

The team at Borexino, a vat containing 300 tonnes of liquid buried under the Italian Alps, captured 14 candidate geoneutrinos between December 2007 and August 2012 (ref. 2). Scientists at KamLAND, with 1,000 tonnes of liquid, say that they detected 116 probable geoneutrinos between March

- Araki, T. et al. *Nature* **436**, 499–503 (2005).
- Bellini, G. et al. Preprint at <http://arxiv.org/abs/1303.2571> (2013).
- The KamLAND Collaboration. Preprint at <http://arxiv.org/abs/1303.4667> (2013).

## ECOLOGY

# Detective work uncovers under-reported overfishing

Excessive catches by Chinese vessels threaten livelihoods and ecosystems in West Africa.

BY CHRISTOPHER PALA

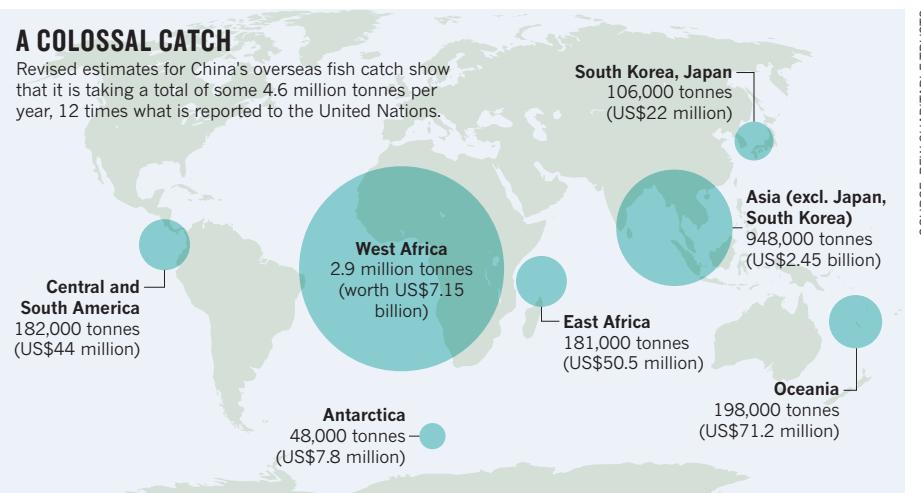
**I**t is a whopper of a catch, in more ways than one: China is under-reporting its overseas fishing catch by more than an order of magnitude, according to a study<sup>1</sup> published on 23 March. The problem is particularly acute in the rich fisheries of West Africa, where a lack of transparency in reporting is threatening efforts to evaluate the ecological health of the waters.

"We can't assess the state of the oceans without knowing what's being taken out of them," says Daniel Pauly, a fisheries scientist at the University of British Columbia in Vancouver, Canada, who led the study. The unreported catch is crippling the artisanal fisheries that help to feed West African populations, he says.

Fisheries experts have long suspected that the catches reported by China to the Food and Agriculture Organization of the United Nations (FAO) in Rome are too low. From 2000 to 2011, the country reported an average overseas catch of 368,000 tonnes a year. Yet China claims to have the world's biggest distant-water fishing fleet, implying a much larger haul, says the study, which was funded by the European Union (EU). Pauly and his colleagues estimate that the average catch for 2000–11 was in fact 4.6 million tonnes a year, more than 12 times the reported figure (see 'A colossal catch'). Of that total, 2.9 million tonnes a year came from West Africa, one of the world's most productive fishing grounds.

Liu Xiaobing, director of the division of international cooperation of China's bureau of fisheries, put the yearly overseas catch at 1.15 million tonnes in a speech to the EU last June. Pauly says that figure would be accurate if it referred to the amount brought back to China, rather than the total catch. Liu did not respond to an e-mailed request for comment.

Fisheries scientists find the latest assessment startling. "So that's where my fish were going!" says Didier Gascuel at the European University of Brittany in Rennes, France, who is a member of the scientific committee that advises Mauritania and the EU on fishing agreements. Year after year, Mauritanian populations of bottom-dwelling species such as octopus, grouper and sea bream have remained stubbornly low — a sign of overfishing by bottom-scraping trawlers, he says. "We had no idea the Chinese catch was so big and of course



SOURCE: PEW CHARITABLE TRUSTS

we never included it our models," he says.

Fishing contracts between Chinese companies and African nations are secret, so to estimate the catch, Pauly and his team had to do some sleuthing. The picture was further clouded because Chinese companies sometimes operate vessels flying local flags. So at least ten researchers combined clues from field interviews, scholarly articles and newspaper and online reports in 14 languages to estimate how many Chinese fishing vessels were operating in 93 countries and territories. They found many in nations where China reported no catch. The estimates were averaged to reach their conclusion: China had at least 900 ocean-going vessels, with 345 in West Africa, including 256 bottom-trawlers.

The scientists estimated the catch per country on the basis of an assumed average catch for each type of vessel. "These numbers may not be absolutely exact, but they give the first hint of the magnitude of the problem," says Boris Worm, a marine ecologist at Dalhousie University in Halifax, Canada, who was not involved in the study.

Other experts are sceptical. "The new estimates seem far, far too high," says Richard Grainger, chief of the fisheries statistics and information service at the FAO. He notes

► NATURE.COM  
For more on  
assessing fish  
stocks, see:  
[go.nature.com/9pjxln](http://go.nature.com/9pjxln)

that a previous assessment<sup>2</sup> estimated the total unreported catch in West Africa (by all nations) at 300,000–560,000 tonnes a year. That study tried

to identify what was missing from official catch figures with a review of English-language scientific studies.

If the new numbers stand up, renewals of fishing contracts with West African nations could be affected. In the 2000s, under public pressure, EU fleets stopped fishing in coastal waters off much of West Africa, except Mauritania and Morocco. They were replaced by Chinese vessels, mostly large bottom-trawlers whose violations of near-shore no-fishing zones have led to protests.

Gascuel, who helps to determine how many fish can be caught while avoiding population collapse, says that numbers of octopus and shrimp available to be taken in EU contracts with Mauritania, primarily by Spanish vessels, were already small. But once the actual Chinese catch is factored in, he says, "we'd have to eliminate the Spanish catch".

Ironically, it was Pauly's team that 12 years ago found that China had been over-reporting its domestic catch by at least 6 million tonnes. Pauly says that mid-level bureaucrats in the country often exaggerate their achievements<sup>3</sup>.

But he says that China's under-reporting of the distant-water catch is the more important problem. "It shows the extent of the looting of Africa, where so many people depend on seafood for basic protein." ■

1. Pauly, D. et al. *Fish Fish.* <http://dx.doi.org/10.1111/faf.12032> (2013).
2. Agnew, D. J. et al. *PLoS ONE* **4**, e4570 (2009).
3. Watson, R. & Pauly, D. *Nature* **414**, 534–536 (2001).



ILLUSTRATION BY ANDY POTTS

**Will an astronaut  
who falls into a black  
hole be crushed or  
burned to a crisp?**

# FIRE IN THE HOLE!

BY ZEEYA MERALI

I

In March 2012, Joseph Polchinski began to contemplate suicide — at least in mathematical form. A string theorist at the Kavli Institute for Theoretical Physics in Santa Barbara, California, Polchinski was pondering what would happen to an astronaut who dived into a black hole. Obviously, he would die. But how?

According to the then-accepted account, he wouldn't feel anything special at first, even when his fall took him through the black hole's event horizon: the invisible boundary beyond which nothing can escape. But eventually — after hours, days or even weeks if the black hole was big enough — he would begin to notice that gravity was tugging at his feet more strongly than at his head. As his plunge carried him inexorably downwards, the difference in forces would quickly increase and rip him apart, before finally crushing his remnants into the black hole's infinitely dense core.

But Polchinski's calculations, carried out with two of his students — Ahmed Almheiri and James Sully — and fellow string theorist Donald Marolf at the University of California, Santa Barbara (UCSB), were telling a different story<sup>1</sup>. In their account, quantum effects would turn the event horizon into a seething maelstrom of particles. Anyone who fell into it would hit a wall of fire and be burned to a crisp in an instant.

The team's verdict, published in July 2012, shocked the physics community. Such firewalls would violate a foundational tenet of physics that was first articulated almost a century ago by Albert Einstein, who used it as the basis of general relativity, his theory of gravity. Known as the equivalence principle, it states in part that an observer falling in a gravitational field — even the powerful one inside a black hole — will see exactly the same phenomena as an observer floating in empty space. Without this principle, Einstein's framework crumbles.

Well aware of the implications of their claim, Polchinski and his co-authors offered an alternative plot ending in which a firewall does not form. But this solution came with a huge price. Physicists would have to sacrifice the other great pillar of their science: quantum mechanics, the theory governing the interactions between subatomic particles.

The result has been a flurry of research papers about firewalls, all struggling to resolve the impasse, none succeeding to everyone's satisfaction. Steve Giddings, a quantum physicist at the UCSB, describes the situation as "a crisis in the foundations of physics that may need a revolution to resolve".

With that thought in mind, black-hole experts came together last month at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, to grapple with the issue face to face. They hoped to reveal the path towards a unified theory of 'quantum gravity' that brings all the fundamental forces of nature under one umbrella — a prize that has eluded physicists for decades.

The firewall idea "shakes the foundations of what most of us believed about black holes", said Raphael Bousso, a string theorist at the University of California, Berkeley, as he opened his talk at the meeting. "It essentially pits quantum mechanics against general relativity, without giving us any clues as to which direction to go next."

#### FIERY ORIGINS

The roots of the firewall crisis go back to 1974, when physicist Stephen Hawking at the University of Cambridge, UK, showed that quantum effects cause black holes to run a temperature<sup>2</sup>. Left in isolation, the holes will slowly spew out thermal radiation — photons and other particles — and gradually lose mass until they evaporate away entirely (see 'The information paradox').

These particles aren't the firewall, however; the subtleties of relativity guarantee that an astronaut falling through the event horizon will not notice this radiation. But Hawking's result was still startling — not least because the equations of general relativity say that black holes can only swallow mass and grow, not evaporate.

Hawking's argument basically comes down to the observation that in the quantum realm, 'empty' space isn't empty. Down at this sub-sub-microscopic level, it is in constant turmoil, with pairs of particles and their corresponding antiparticles continually popping into existence before rapidly recombining and vanishing. Only in very delicate laboratory experiments does this submicroscopic frenzy have any observable consequences. But when a particle–antiparticle pair appears just outside a black hole's event horizon,

Hawking realized, one member could fall in before the two recombined, leaving the surviving partner to fly outwards as radiation. The doomed particle would balance the positive energy of the outgoing particle by carrying negative energy inwards — something allowed by quantum rules. That negative energy would then get subtracted from the black hole's mass, causing the hole to shrink.

Hawking's original analysis has since been refined and extended by many researchers, and his conclusion is now accepted almost universally. But

## "IT IS A CRISIS IN THE FOUNDATIONS OF PHYSICS."

with it came  
the disturbing realization that  
black-hole radiation leads to a paradox that chal-  
lenges quantum theory.

Quantum mechanics says that information cannot be destroyed. In principle, it should be possible to recover everything there is to know about the objects that fell in a black hole by measuring the quantum state of the radiation coming out. But Hawking showed that it was not that simple: the radiation coming out is random. Toss in a kilogram of rock or a kilogram of computer chips and the result will be the same. Watch the black hole even until it dies, and there would still be no way to tell how it was formed or what fell in it.

This problem, dubbed the black-hole information paradox, divided physicists into two camps. Some, like Hawking, argued that the information truly vanishes when the black hole dies. If that contradicted quantum laws, then better laws needed to be found. Others, like John Preskill, a quantum physicist at the California Institute of Technology in Pasadena, stuck by quantum mechanics. "For a time, I did seriously try to build an alternative theory that included information loss," he says. "But I couldn't find one that made any sense — nobody could." The stalemate continued for the next two decades, finding its most famous expression in 1997, when Preskill publicly bet Hawking that information was not being lost, with the winner to receive an encyclopaedia of his choice.

But that same year, the deadlock was broken by a discovery made by Juan Maldacena, a physicist then at Harvard University in Cambridge. Maldacena's insight built on an earlier proposal that any three-dimensional (3D) region of our Universe can be described by information encoded on its two-dimensional (2D) boundary<sup>3–5</sup>, in much the same way that laser light can encode a 3D scene on a 2D hologram. "We used the word 'hologram' as a metaphor," says Leonard Susskind, a string theorist at Stanford University in California, and one of those who came up with the proposal<sup>4</sup>. "But after doing more mathematics, it seemed to make literal sense that the Universe is a projection of information on the boundary."

What Maldacena came up with was a concrete mathematical

formulation<sup>6</sup> of the hologram idea that made use of ideas from superstring theory, which posits that elementary particles are composed of tiny vibrating loops of energy. His model envisages a 3D universe containing strings and black holes that are governed only by gravity, bounded by a 2D surface on which elementary particles and fields obey ordinary quantum laws without gravity. Hypothetical residents of the 3D space would never see this boundary because it is infinitely far away. But that wouldn't matter: anything happening in the 3D universe could be described equally well by equations in the 2D universe, and vice versa. "I found that there's a mathematical dictionary that allows you to go back and forth between the languages of these two worlds," Maldacena explains.

This meant that even 3D black-hole evaporation could be described in the 2D world, where there is no gravity, where quantum laws reign supreme and where information can never be lost. And if information is preserved there, then it must also be preserved in the 3D world. Somehow, information must be escaping from the black holes.

#### ONE FOR ALL

A few years later, Marolf showed that every model of quantum gravity will obey the same rules, whether or not it is built from string theory<sup>7</sup>. "It was a combination of Maldacena and Marolf's work that turned me around," explains a long-term proponent of information loss, Ted Jacobson, a quantum physicist at the University of Maryland in Baltimore. In 2004, Hawking publicly admitted that he had been wrong, and gave

For a particle to be emitted at all, it has to be entangled with the twin that is sacrificed to the black hole. And if Susskind and others were right, it also had to be entangled with all the Hawking radiation emitted before it. Yet a rigorous result of quantum mechanics dubbed 'the monogamy of entanglement' says that one quantum system cannot be fully entangled with two independent systems at once.

To escape this paradox, Polchinski and his co-workers realized, one of the entanglement relationships had to be severed. Reluctant to abandon the one required to encode information in the Hawking radiation, they decided to snip the link binding an escaping Hawking particle to its infalling twin. But there was a cost. "It's a violent process, like breaking the bonds of a molecule, and it releases energy," says Polchinski. The energy generated by severing lots of twins would be enormous. "The event horizon would literally be a ring of fire that burns anyone falling through," he says. And that, in turn, violates the equivalence principle and its assertion that free-fall should feel the same as floating in empty space — impossible when the former ends in incineration. So they posted a paper on the pre-print server, arXiv, presenting physicists with a stark choice: either accept that firewalls exist and that general relativity breaks down, or accept that information is lost in black holes and quantum mechanics is wrong<sup>1</sup>. "For us, firewalls seem like the least crazy option, given that choice," says Marolf.

The paper rocked the physics community. "It was outrageous to claim that giving up Einstein's equivalence principle is the best option," says Jacobson. Bousso agrees, adding: "A firewall simply can't appear in empty space, any more than a brick wall can suddenly appear in an empty field and smack you in the face." If Einstein's theory doesn't apply at the event horizon, cosmologists would have to question whether it fully applies anywhere.

Polchinski admits that he thought they could have made a silly mistake. So he turned to Susskind, one of the fathers of holography, to find it. "My first reaction was that they were wrong," says Susskind.

He posted a paper stating as much<sup>8</sup>, before quickly retracting it, after further thought. "My second reaction was that they were right, my third was that they were wrong again, my fourth was that they were right," he laughs. "It's earned me the nickname, 'the yo-yo,' but my reaction is pretty much the same as most physicists."

Since then, more than 40 papers have been posted on the topic in arXiv, but as yet, nobody has found a flaw in the team's logic. "It's a really beautiful argument proving that there's something inconsistent in our thinking about black holes," says Don Page, a collaborator of Hawking's during the 1970s who is now at the University of Alberta in Edmonton, Canada. A number of inventive solutions have been offered, however.

#### REAL-WORLD IMPLICATIONS

One of the most promising resolutions, according to Susskind, has come from Daniel Harlow, a quantum physicist at Princeton University in New Jersey, and Patrick Hayden, a computer scientist at McGill University in Montreal, Canada. They considered whether an astronaut could ever detect the paradox with a real-world measurement. To do so, he or she would first have to decode a significant portion of the outgoing Hawking radiation, then dive into the black hole to examine the infalling particles. The pair's calculations show that the radiation is so tough to decode that the black hole would evaporate before the astronaut was ready to jump in<sup>9</sup>. "There's

# "THE EVENT HORIZON IS LITERALLY A RING OF FIRE."

Preskill a baseball encyclopaedia to make good on their bet.

Such was the strength of Maldacena's discovery that most physicists believed that the paradox had been settled — even though nobody had yet explained how Hawking radiation smuggles information out of the black hole. "I guess we just all assumed there would be a straightforward answer," says Polchinski.

There wasn't. When Polchinski and his team set themselves the task of clearing up that loose end in early 2012, they soon stumbled on yet another paradox — the one that eventually led them to the fatal firewall.

Hawking had shown that the quantum state of any one particle escaping from the black hole is random, so the particle cannot be carrying any useful information. But in the mid-1990s, Susskind and others realized that information could be encoded in the quantum state of the radiation as a whole if the particles could somehow have their states 'entangled' — intertwined in such a way that measurements carried out on one will immediately influence its partner, no matter how far apart they are.

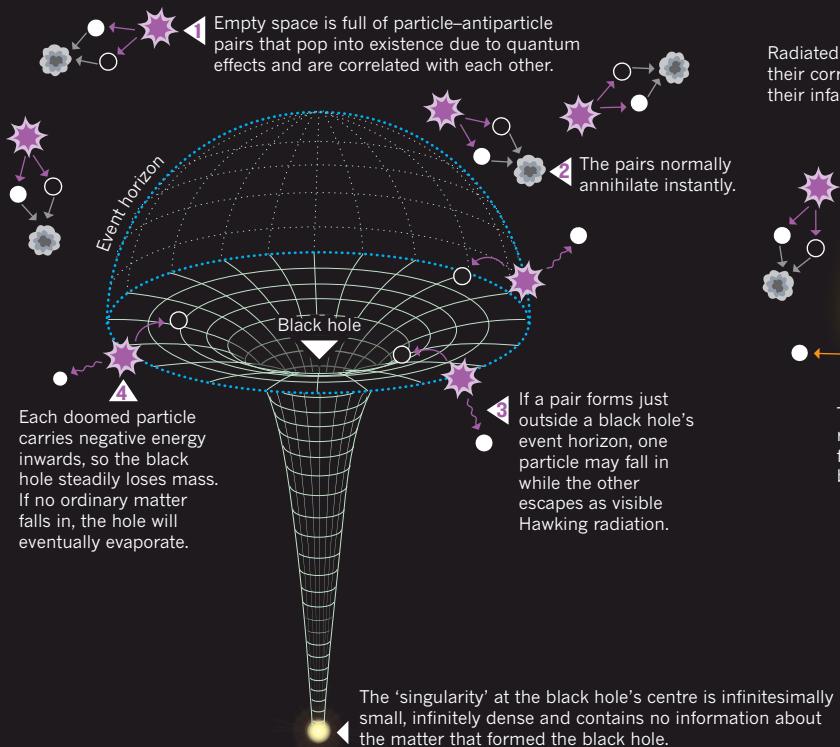
But how could that be, wondered the Polchinski's team?

# THE INFORMATION PARADOX

Matter falling into a black hole is crushed to an infinitely dense point at the centre. Two scenarios attempt to explain what happens to the information that matter holds.

## DISAPPEARANCE

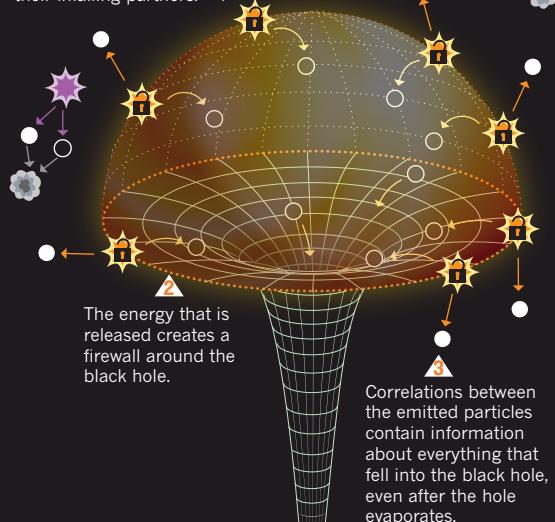
When the hole evaporates, all the information disappears with it.



## FIREWALL

The information is carried out by quantum correlations among all particles radiated from the black hole.

Radiated particles break their correlations with their infalling partners.



no fundamental law preventing someone from measuring the paradox," says Harlow. "But in practice, it's impossible."

Giddings, however, argues that the firewall paradox requires a radical solution. He has calculated that if the entanglement between the outgoing Hawking radiation and its infalling twin is not broken until the escaping particle has travelled a short distance away from the event horizon, then the energy released would be much less ferocious, and no firewall would be generated<sup>10</sup>. This protects the equivalence principle, but requires some quantum laws to be modified. At the CERN meeting, participants were tantalized by the possibility that Giddings' model could be tested: it predicts that when two black holes merge, they may produce distinctive ripples in space-time that can be detected by gravitational-wave observatories on Earth.

There is another option that would save the equivalence principle, but it is so controversial that few dare to champion it: maybe Hawking was right all those years ago and information is lost in black holes. Ironically, it is Preskill, the man who bet against Hawking's claim, who raised this alternative, at a workshop on firewalls at Stanford at the end of last year. "It's surprising that people are not seriously thinking about this possibility because it doesn't seem any crazier than firewalls," he says — although he adds that his instinct is still that information survives.

The reluctance to revisit Hawking's old argument is a sign of the immense respect that physicists have for Maldacena's dictionary relating gravity to quantum theory, which seemingly proved that information cannot be lost. "This is the deepest ever insight into gravity because it links it to quantum fields," says Polchinski, who compares Maldacena's

result — which has now accumulated close to 9,000 citations — to the nineteenth-century discovery that a single theory connects light, electricity and magnetism. "If the firewall argument had been made in the early 1990s, I think it would have been a powerful argument for information loss," says Bousso. "But now nobody wants to entertain the possibility that Maldacena is wrong."

Maldacena is flattered that most physicists would back him in a straight-out fight against Einstein, although he believes it won't come to that. "To completely understand the firewall paradox, we may need to flesh out that dictionary," he says, "but we won't need to throw it out."

The only consensus so far is that this problem will not go away any time soon. During his talk, Polchinski fielded all proposed strategies for mitigating the firewall, carefully highlighting what he sees as their weaknesses. "I'm sorry that no one has gotten rid of the firewall," he concludes. "But please keep trying." ■

**Zeeya Merali** is a freelance writer based in London.

1. Almheiri, A., Marolf, D., Polchinski, J. & Sully, J. Preprint at <http://arxiv.org/abs/1207.3123> (2012).
2. Hawking, S. W. *Nature* **248**, 30–31 (1974).
3. Bekenstein, J. D. *Phys. Rev. D* **7**, 2333–2346 (1973).
4. Susskind, L. *J. Math. Phys.* **36**, 6377 (1995).
5. Stephens, C. R., 't Hooft, G. & Whiting, B. F. *Class. Quant. Grav.* **11**, 621–647 (1994).
6. Maldacena, J. M. *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).
7. Marolf, D. *Phys. Rev. D* **79**, 044010 (2009).
8. Susskind, L. Preprint at <http://arxiv.org/abs/1207.4090> (2012).
9. Harlow, D. & Hayden, P. Preprint at <http://arxiv.org/abs/1301.4504> (2013).
10. Giddings, S. B. Preprint at <http://arxiv.org/abs/1302.2613> (2013).

# COMMENT

**COLLABORATIONS** Choosing the right people to work across disciplines **p.27**



**PALAEONTOLOGY** A tour of excavation sites charts the rise and fall of dinosaurs **p.30**

**WOMEN IN SCIENCE** A selection of responses to *Nature's* special on gender equality **p.31**

**OBITUARY** Donald Arthur Glaser, the inventor of the bubble chamber **p.32**

DIMA KOROTAYEV/REUTERS



A Russian defence unit practises rescue training after a simulated chemical attack in 2002.

## Update the Chemical Weapons Convention

Bring biological threats into the treaty and make chemists more aware of the dark side of their research, says **Leiv K. Sydnes**.

From 8–19 April, representatives of the 188 nations that signed and ratified the Chemical Weapons Convention (CWC) will meet in The Hague, the Netherlands, to review the treaty for the third time. The documents at the heart of the discussions bear good news about progress against the threat of chemical warfare. But they do not address recent geopolitical changes or emerging small-scale production technologies that pose new risks.

Analyses from the past five years will show that the Organisation for the Prohibition of

Chemical Weapons (OPCW) has done an outstanding job. About 78% of the declared chemical-weapons stockpile has been destroyed, and this is expected to rise to 99% by 2017. The reports will also show that all known chemical-weapons production facilities have been deactivated or converted for peaceful purposes. And they will state that scientific and technological developments have been evolutionary, not dramatic, with little impact on the production and use of chemical weapons.

But what will be missing is a crisp

assessment of how major geopolitical changes have altered the world that the CWC and OPCW were created to serve. The CWC is an agreement between states, as these were the political players of the past. But major conflicts today involve a few countries that have not signed the convention, and some small, militant groups with political objectives but no legal standing. Syria, for example — one of six countries that has neither signed nor acceded to the CWC — is said to have chemical weapons. This undermines the authority of the convention. ▶

► New chemical processing techniques and equipment, including compact reactors, mean that terrorists can make small but deadly quantities of chemical weapons beyond the reach of the convention. Toxic compounds can also be synthesized using biological processes, which are not controlled.

These facts call for two science-based actions: inspection of biotech industries, and thorough analysis of data showing sales of chemicals and reactor technology. To achieve this, we should rethink the CWC and merge it with the Biological and Toxin Weapons Convention (BWC). An updated convention needs a broader remit if it is to limit the re-emergence of chemical weapons in any form and any amount, anywhere.

#### LIMITED SCOPE

Conferences take place every five years to evaluate the operation of the CWC treaty, assess progress and decide whether modifications are needed. So far, few changes have been made. The signatory countries have been concerned with chemical warfare on a massive scale, such as the Iran–Iraq war (1980–88). The regulations and inspection regimes focus on certain types of chemicals owned by states in large quantities — both chemical weapons, such as sarin and mustard gas, and the chemicals, including phosgene (carbonyl dichloride) and thionyl chloride, that are used to prepare them.

Geopolitical changes affect the CWC because it deals only with national authorities and not with the powerful individuals, revolutionary groups, factions in civil wars and terrorist cells that can exert a detrimental influence on global security. The number of non-state political actors has grown as a result of regional conflicts, increasing terrorism and the collapse of national structures. The CWC does not give the OPCW any authority to act in conflicts between such parties. This does not prevent the destruction of existing chemical weapons, but it could pose serious problems in the future.

The small-scale production of chemical weapons by activists and terrorists is also beyond the scope of the CWC, and these risks are growing. Advances in three areas must be watched carefully.

First, closer surveillance is needed of chemistry that overlaps with medicine and biology, such as research on neuropeptides and bioregulators — an area that is growing rapidly and paves the way for synthetic biology and the large-scale production of toxins. Examples are saxitoxin (a paralytic



Industrial plants in Syria are outside the remit of the Chemical Weapons Convention.

neurotoxin) and ricin (an inhibitor of protein production), which are prohibited by both the BWC and the CWC. It therefore makes sense to merge the two conventions and introduce the CWC inspection regime to the biological sciences. This will have a significant impact because the BWC does not authorize inspections.

A second area for scrutiny is research on incapacitating chemical agents (ICAs) — compounds that act on the central nervous system but are meant for crowd control, rather than warfare<sup>1</sup>. Such work is currently allowed under the CWC, but I agree with international security expert Malcolm Dando of the University of Bradford, UK, that it should be prohibited<sup>2</sup>. Compounds such as fentanyl derivatives and other opioids are deadly to some people but not to others, depending on gender, age and general health. When some 750 hostages were exposed to a fentanyl derivative in a Moscow theatre in 2002, for example, about 125 people died. Other deadly ICA compounds will inevitably be produced in the search for a safe agent for riot control.

Third, flow microreactors must be tracked because they make it relatively safe to produce toxic compounds, as only small amounts of products are present in the reactor at any given time. Microreactors have become widespread because they are robust and easy to use, making it quick and easy to set up and dismantle a small chemical-weapons facility. The OPCW should consider holding a register of microreactor buyers.

Information and people are both spread more widely than five years ago. The scientific literature is accessible anywhere with adequate computer facilities, and can be searched quickly and thoroughly. With procedures and

starting materials so readily available, almost anyone can try to make a chemical weapon.

Many students are educated in several different countries and become involved in multinational research projects. Researchers travel to lecture, to discuss and build networks, and to participate in virtual education. This mobility increases the likelihood that individuals with a university education in chemistry will be exposed to political ideas that lead them to commit terrorist acts. Indeed, at least one of the terrorists behind the London bombing in 2005 had some education in chemistry.

#### CODE OF CONDUCT

Chemistry cannot be blamed for chemical warfare or terrorism, but the chemistry community has a duty to be aware

of the danger and to act to prevent the misuse of chemicals. We must work to educate people about chemical safety, waste disposal and the responsible use of chemicals, starting at school. We need university courses on chemical weapons, and I welcome initiatives by the OPCW and several non-governmental organizations, including the International Union of Pure and Applied Chemistry, to provide materials for these.

Awareness also has an ethical dimension. The OPCW is among those who have tried to compose a code of conduct for chemists, but so far no document has emerged. Some say that the civilian and military dual use of chemicals makes it almost impossible to agree on a short and precise wording. Others argue that signing a code means nothing if you are inclined to violate the CWC.

I would like to see a code of conduct. I think that signing up to such a code would force chemists to reflect on the gravity of their work, increasing their levels of responsibility and awareness. This, a strong OPCW and a new Biological and Chemical Weapons Convention are the best ways to prevent the use of these terrible weapons. ■

**Leiv K. Sydnes** is professor in the Department of Chemistry, University of Bergen, NO-5007 Bergen, Norway. He chaired the international task group that assessed the impact of scientific advances on the CWC in 2007 and 2012.  
e-mail: leiv.sydnes@kj.uib.no

- Mogl, S. (ed.) *Technical Workshop on Incapacitating Chemical Agents: Spiez, Switzerland, 8–9 September 2011* (Spiez Laboratory, 2012); available at go.nature.com/eqsqoh.
- Dando, M. *Nature* **460**, 950–951 (2009).

# Link the world's best investigators

Nurturing small groups of leading researchers – especially young scientists – is the way to break intellectual ground, says **Alan Bernstein**.

**W**hen the molecular-biology revolution began in the 1950s, biochemist Erwin Chargaff observed that the physicists, chemists and geneticists involved were “practising biochemistry without a licence”. He didn’t recognize that this small group of young innovators was forging an entirely new approach to biology.

Many challenges facing humanity, from chronic disease to climate change, demand a multidisciplinary perspective and fresh thinking — as well as a global approach. Yet few seasoned researchers venture beyond their scientific fields, campuses or national funding programmes. The obstacles are even greater for young scientists, whose creativity is constrained within large teams dominated by senior investigators, and whose prospects for funding and positions are dwindling.

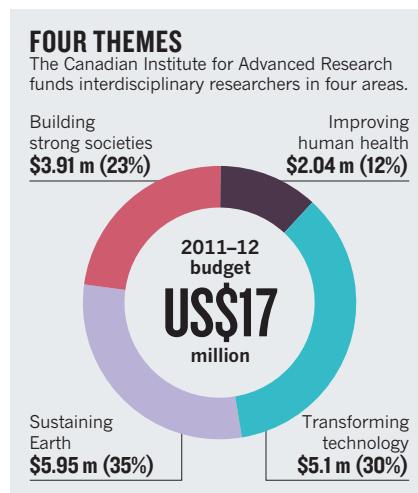
As I approach the end of my first year as president of the Canadian Institute for Advanced Research (CIFAR), I have been reflecting on how best to support transformative advances in science. As globalization and the pace and scale of research increase, different sorts of institutions and funding models are needed. As a virtual global network, CIFAR offers one vision for supporting exchanges across borders and disciplines. But more needs to be done to enable talented young scientists to pursue original, multidisciplinary work.

## GLOBAL COLLABORATION

For three decades, CIFAR has brought together small groups of outstanding researchers from varied fields to focus on important scientific questions. Rather than replicating the university campus, as other advanced institutes do, CIFAR functions as a virtual academy. From its headquarters in Toronto, CIFAR’s 12 networks — which cover topics from human health and quantum materials to economics and sustainability — link almost 400 fellows and advisers from 103 institutions in 16 countries (see ‘Four themes’). Landmark studies so far include *Why Are Some People Healthy and Others Not?*<sup>1</sup> and *Why Nations Fail*<sup>2</sup>.

CIFAR is not a granting agency, but convenes meetings for its global programmes, typically twice a year. With an operating budget in 2013 of Can\$17 million (US\$17 million), CIFAR is funded by federal and provincial Canadian governments, individuals and the private sector.

The institute’s recipe for success has three key ingredients: framing the right question,



identifying the right researchers and leader, and supporting a network for long enough to overcome disciplinary and cultural barriers.

The compelling questions that interest CIFAR should presage a new field of enquiry, require multiple perspectives to answer and have enough appeal to bond a diverse group. The question of how early social experiences affect an individual’s life-long physical and mental health and well-being, for example, is the focus of our Child & Brain Development programme<sup>3</sup>. This year, CIFAR will launch its first Global Call for Ideas, inviting the Canadian and world research community to propose new questions on global challenges.

Each programme is overseen by an external advisory committee of 3–5 distinguished investigators who help to select members and direct the research. Programmes are reviewed every 5 years by an external panel on the basis of scientific excellence, synergy and impact. Programmes may be terminated, extended for a further 5 years, or renewed with different leadership or in a new direction.

The choice of members is crucial. Scientific excellence is a prerequisite, and ‘talent attracts talent’. The programmes’ disciplinary and geographic diversity make it essential that members have exemplary communication skills. Leadership also matters, so programme directors must have high standing and experience of running their own groups. To develop novel ideas, they also need listening skills and the ability to synthesize viewpoints.

It takes time to find common ground, so CIFAR typically supports programmes for 10–15 years. Deep dialogue requires trust, which is built through face-to-face exchanges. The use of small groups (15–20 people)

retains intimacy while encouraging robust discussions and risk taking.

There are no requirements for how much time each fellow must devote to the programme; collaborations arise naturally. Most fellows say that their involvement changes the long-term direction of their research.

Some might argue that the hand-picking of fellows is elitist, but I believe that it provides an opportunity for the best to succeed. On the advice of the programme director and the advisory committee, fellows are approved by the president, with appointments monitored closely by CIFAR’s board of directors to avoid bias. Nevertheless, the involvement of one group still needs to be strengthened: that of young investigators.

## THE NEXT GENERATION

Young investigators are crucial to the future of research, especially at a time of profound technological change. Computational biology, for example, is largely driven by emerging scientists. But young scientists face many hurdles, from low grant success rates to a dearth of academic positions and the difficulty of developing autonomy within large teams. Those in the developing world are further hindered by scarce resources, remoteness, and heavy teaching and administrative loads<sup>4</sup>.

Although CIFAR has brought 24 outstanding young researchers into its existing programmes, we recognize that this is not enough. In February, we launched the CIFAR Global Academy, an initiative that will help emerging researchers to develop their careers. It will include visits and exchanges, meetings and programmes at the interface between research, policy and practice.

We need to increase the world’s capacity to collaborate and to nurture the next generation of research leaders. We hope that other organizations will work with us to explore new partnerships and models of interaction across disciplines, nations and generations. ■

**Alan Bernstein** is president of the Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, Canada.  
e-mail: abernstein@cifar.ca

- Evans, R. G., Barer, M. L. & Marmor, T. R. *Why Are Some People Healthy and Others Not?* (Walter de Gruyter, 1994).
- Acemoglu, D. & Robinson, J. A. *Why Nations Fail* (CROWN Business, 2012).
- Hertzman, C. *Proc. Natl Acad. Sci. USA* **109** (suppl. 2), 17160–17167 (2012).
- Barouch, D. H. et al. *Nature Precedings* <http://dx.doi.org/10.1038/npre.2010.4800.2> (2010).



Waura tribesmen fish in the Piulaga lake near their village in the Upper Xingu region of Brazil's Mato Grosso state.

ANTHROPOLOGY

# A monochrome Eden

**Bob Bloomfield** assesses a haunting photographic record of remote environments and the indigenous peoples who live in them.

The celebrated Brazilian documentary photographer Sebastião Salgado is best known for his black-and-white images of the Serra Pelada gold mine in Brazil, shot in 1986. Those stark photographs of mud-encrusted workers toiling through the labyrinthine open mine workings highlighted the appalling working conditions. Now, after decades of social observation, Salgado has turned his lens on the environment, to convey how its degradation has led to much of the inequity he has recorded.

*Genesis* is a project eight years in the making — the culmination of journeys across 32 countries to document remote ecosystems and local peoples whose survival now hangs by a thread. The 250 photographs on show at London's Natural History Museum are part of a portfolio that is touring venues around the world. Salgado calls his new work *Genesis* because he wanted to record “the animal species that have resisted domestication” and the “remote tribes whose ‘primitive’ way of life is largely untouched”. He hopes that by highlighting this “uncontaminated world” he can help it to be preserved and, where degraded,

restored. His vision is that development need not be synonymous with destruction.

The images include breathtaking vistas across Siberian wastes, in verdant Amazonian forested valleys, over the haunting Namibian dune fields and among the animals at home in the harsh beauty of Antarctica. He spent time with local peoples, including the Xingu in the upper Amazon, the Bushmen in Botswana and the Dinkas in Sudan. There are striking records of daily life: foraging, making fires, preparing food, socializing and sharing ritual celebrations.

Salgado provides no contrasting images of the globalized world, dominated by urban expansion and the conversion of natural land for burgeoning infrastructure, transportation, extraction and agro-forestry. Nor, in hymning the intrinsic beauty of wilderness, does he draw attention to those regions' wider value as vital ‘organs of the biosphere’, contributing to the planet's regulatory system. There is

**Sebastião Salgado: Genesis**  
WATERHOUSE GALLERY, NATURAL HISTORY MUSEUM, LONDON  
11 April to 8 September 2013

instead a nostalgia for a halcyon disappearing world — worthy and understandable, but saying little about the broader imperative to protect global wilderness and biodiversity.

Similar issues cloud the portraits. Salgado spent time with the people he recorded, and the images were taken with their consent. Yet his photographs of proud people seem to be coloured by the Arcadian ideals of the late eighteenth century — a vision of ‘noble savages’ populating an Eden somehow preserved from the ravages of civilization. In part because of the likeness of the black-and-white images to old sepia photographs, I found myself reminded of colonial and ethnographic photographic archives.

Today, our understanding of indigenous peoples is more nuanced. Even though they may be profoundly culturally adapted to their environments, their ‘sustainable’ lifestyles often come at a price, such as high child mortality or relatively short lifespans. And whereas many indigenous cultures are seriously threatened by outside forces, many are increasingly proactive in fighting for their traditional and customary rights

SEBASTIÃO SALGADO / AMAZONAS IMAGES/NBPICTURES

to the land, and are adopting new ideas and technologies on their own terms.

For example, in Brazil, the Kaxinawá live in the state of Acre in the western Amazon, where forest covers 88% of the land but indigenous groups control only 1.5% of it. Kaxinawá elders are working with the state government to ensure official recognition of and compensation for their traditional stewardship of the forests. In Ecuador, the Huaorani, whose lands are under threat from logging and oil interests, are using geographic information system technology and traditional knowledge to map areas for potential tourism and conservation, creating a community resource for planning and negotiating. There is nothing of this side of the story in *Genesis*.

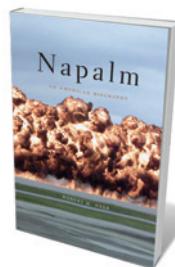
Salgado spends time on the Atlantic coast of Brazil, home to one of the most biodiverse forest ecosystems in the world, the highly threatened Atlantic Forest (Mata Atlântica), of which only small fragments survive. With his wife, Lélia Deluiz Wanick, Salgado has been restoring a denuded 630-hectare site near Aimorés by planting trees and reintroducing other species. They have also founded the Instituto Terra, an initiative that supports the environmental education of Brazilian officials, teachers and farmers, and aims to show that it is possible to restore environments that seemed lost forever.

Although *Genesis* brings into sharp focus the need to preserve both the wilderness and the well-being of the peoples relying on it, the role of the outside world is more ambiguous. Both the Atlantic Forest project and the exhibition are supported by the Brazilian multinational Vale, based in Rio de Janeiro, one of the biggest construction and mining companies in the world. Its website states that "it is committed to quality of life and environmental preservation". However, some indigenous peoples and human-rights groups oppose Vale's stakes in major dam and mining projects in the Amazon interior.

If *Genesis* gets people to reconsider the value of the disappearing wild, it will have achieved an important aim. If it promotes discussion on the plight of indigenous peoples, their self-determination and their potential as stewards of wilderness crucial to our future, it will have offered an invaluable revelation. But however visually stunning and moving these images are, to me they reduce the complex realities for local people and remote environments of our changing world to black and white, rather than shades of grey. ■

**Bob Bloomfield** is an independent expert, based in the United Kingdom, on public engagement and understanding of science and environment, and formerly in the Public Engagement Executive of London's Natural History Museum.  
e-mail: bob\_bloomfield@bobania.f9.co.uk

# Books in brief



## Napalm: An American Biography

Robert M. Neer HARVARD UNIVERSITY PRESS 352 pp. \$29.95 (2013)

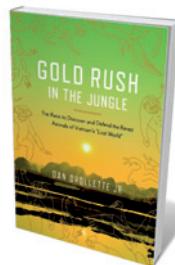
In 1942, in a secret lab at Harvard University in Massachusetts, chemist Louis Fieser and his team created napalm — an incendiary gel that sticks to skin and can burn down to the bone. Robert Neer's harrowing story veers from Fieser's tests on 'kamikaze' bats fitted with napalm bombs to the 1944–45 incendiary bombing of Japan that killed 330,000 people. In 1980, the United Nations declared the gel's use on concentrations of civilians a war crime. An interview with Phan Thi Kim Phuc — photographed running naked and napalm-burned during the Vietnam War — underlines the cost of miscast innovation.



## Invisible in the Storm: The Role of Mathematics in Understanding Weather

Ian Roulstone and John Norbury PRINCETON UNIVERSITY PRESS 346 pp. \$35 (2013)

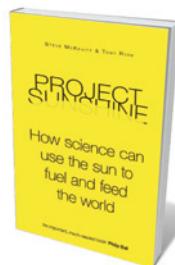
Mathematicians Ian Roulstone and John Norbury demystify the maths behind meteorology. Trailblazers' work is vividly evoked, from eighteenth-century mathematician Leonhard Euler on hydrostatics to physicist Vilhelm Bjerknes's numerical weather prediction. The pace cranks up with twentieth-century advances such as Jule Gregory Charney's harnessing of the gargantuan ENIAC computer for his work in the 1940s and 1950s on forecasting pressure patterns.



## Gold Rush in the Jungle: The Race to Discover and Defend the Rarest Animals of Vietnam's "Lost World"

Dan Drollette Jr CROWN 336 pp. \$25 (2013)

From elusive forest oxen to barking deer, Vietnam has seen a raft of newly discovered species emerge during the past decade. Almost as swiftly, the black market is bringing many close to extinction. Science journalist Dan Drollette Jr reveals the courage and ingenuity of researchers intent on preserving what wildlife they can. Drawing on years of visits to the country, his snapshots of these wildlife warriors — such as langur specialist Tilo Nadler — reveal approaches that could show the way for conservationists in other tight spots.



## Project Sunshine: How Science Can Use the Sun to Fuel and Feed the World

Tony Ryan and Steve McEvitt ICON 320 pp. £16.99 (2013)

Polymer chemist Tony Ryan and consumerism expert Steve McEvitt train a ray of sunshine on the issues of food and energy security. They argue — and are backed by a major research project at the University of Sheffield, UK — that a "solar economy" can sustainably support the world population of 9 billion people predicted for 2050. Cogently analysing business as usual won't work; the authors lay out a science-based, multi-pronged solution incorporating measures such as state-of-the-art photovoltaics.



## Bug Music: How Insects Gave Us Rhythm and Noise

David Rothenberg ST MARTIN'S PRESS 288 pp. \$26.99 (2013)

Philosopher and musician David Rothenberg asks whether the tapping, clicking, humming leitmotifs of the insect world inspired human rhythm. The result mixes research on insect behaviour and anatomy, sonograms and more with apt artistic digressions. From the deafening 'surround-sound' of the 17-year cicada *Magicicada cassini* to syncing crickets, vibrating three-humped treehoppers and the "penile music" of water boatmen, this is an enchanting foray into the "polyrhythmic swirls of the entomological soundscape".

# Giants unearthed

Xu Xing revels in an enthusiast's tour of the Mesozoic era and its denizens.

**B**rian Switek is a science writer with a deep understanding and broad knowledge of evolution. His books offer a compelling mixture of reliable information, personal experiences and thoughts, anecdotes about palaeontological research and even science philosophy, delivered in a breezy and engaging style.

In *My Beloved Brontosaurus*, Switek — a fan of dinosaurs since childhood — describes his tours of famous excavation sites, selected museums and high-tech research laboratories across the United States. He meets and talks to a host of palaeontologists, including fossil-bone expert Mark Goodwin at the University of California Museum of Paleontology at Berkeley. By getting close to the people and the science that they move forward, Switek paints a comprehensive picture of how our understanding of dinosaur evolution gradually advances. Along the way, he shows how dinosaur palaeontology is a cultural touchstone in the United States and many other countries.

Switek touches on widely discussed topics, such as how dinosaurs rose to dominance early in the Mesozoic era (250 million to 65 million years ago) and why all of them — except the strange feathered ones we know as birds — died out when the era ended. But his focus is on how dinosaurs lived.

Switek picks through recently discovered fossils such as the carnivorous feathered dinosaur *Scirurumimus albersdoerferi* found in Germany in 2011. And he discusses cutting-edge methods such as the use of computed tomography scanning to investigate fossils' nasal passages, brain cavities and other internal cranial features. A recurring theme of the book is that new specimens, technologies and thinking are making it possible to investigate how dinosaurs mated, matured and became gigantic; to study their social life and the diseases that afflicted some individuals; and even, as in the case of the small, bird-like *Anchiornis*, to reconstruct aspects of their coloration.

Some of Switek's best palaeontological stories come from the history of the



Sculptor Brian Cooley's reconstruction of a *Tyrannosaurus rex*.



**My Beloved Brontosaurus: On the Road with Old Bones, New Science, and Our Favorite Dinosaurs**  
BRIAN SWITEK  
*Scientific American*: 2013. 272 pp. \$26

discipline. He explains how his favourite dinosaur when he was a child was popularly known as brontosaurus, but had in fact been known scientifically as *Apatosaurus* since 1903, when it was reclassified as a species of the latter genus. Eventually, the new name filtered through to museum exhibits and popular books, and although brontosaurus may not have disappeared entirely, the current generation of dinosaur-savvy children are aware of its obsolescence.

More often, Switek focuses on recent advances. The rediscovery of an animal called *Effigia* in an old collection of fossils at the American Museum of Natural History in New York helped researchers to show that some early relatives of crocodilians were similar to many dinosaurs — not only did they have a near-vertical rather than sprawling

hindlimb posture, but they were also bipedal. A reluctant field decision to break the thigh bone of a *Tyrannosaurus rex* for ease of transport exposed a type of bone tissue seen today in female birds that are about to lay eggs, revealing the dinosaur's gender. And a graduate student's thesis research on the ink sacs of a fossil squid led to a method for partially reconstructing the colours of some extinct animals, including dinosaurs.

I have a few quibbles. In some instances, Switek recounts a conversation he had with a scientist to lend authority to an account of a discovery, but the details on how inferences were made and hypotheses tested are left out. The phylogeny-based definitions given for some animal groups are slightly inaccurate, and there are occasional lapses in the reliability of the scientific storytelling. For example, in discussing the importance of *Effigia*, Switek glosses over the fact that near-vertical hindlimb posture (although not bipedality) was already known to be widespread among early members of the crocodilian lineage.

The book largely revolves around the dinosaur stars from the early days of North American palaeontology, and space for recent discoveries made by researchers elsewhere is limited. Similarly, some fascinating new research directions are discussed briefly or not at all.

Switek does not mention the 34 juveniles of the small, parrot-beaked dinosaur *Psittacosaurus* that were found preserved with an adult in Chinese strata — a fossil that is among the best evidence to support the existence of parenting in dinosaurs. He also omits the use of isotopic analysis to reconstruct the lifestyle of the mysterious spinosaurs — massive carnivorous dinosaurs with vaguely crocodile-like heads. Nor is there anything about molecular studies on the development of feathers, fingers and toes in modern taxa, which have contributed greatly to our understanding of the morphological evolution of birds.

Of course, not even a brontosaurus-sized book could include everything. Switek has succeeded in covering a wide range of interesting topics in dinosaur palaeontology with infectious enthusiasm. ■

**Xu Xing** is professor in the Institute of Vertebrate Paleontology and Paleoanthropology of the Chinese Academy of Sciences in Beijing.  
e-mail: xu.xing@ivpp.ac.cn

# Correspondence

## Sexism: dearth of female role models

Although "science remains institutionally sexist" (*Nature* **495**, 21; 2013), gender inequality does not operate on its own. It is mediated by other inequalities, such as social class and race, which intersect.

Of the four researchers you use to illustrate the contribution of young, driven women to the scientific landscape (*Nature* **495**, 28–31; 2013), at least two come from families in which both parents were scientists. This could well have habituated them in scientific practice from an early age. The reality is that most female scientists-in-training struggle to find female role models to smooth their transition into an academic career.

This effect is particularly isolating for women of African or Caribbean origin, who are among the most under-represented groups in the sciences.

We call for rigorous qualitative research into the interaction of all these factors so that the

occupational field can be cleared of cultural disadvantages for aspiring women and minority scientists.

**Simon Williams, Christine Wood, Richard McGee Feinberg**  
School of Medicine, Northwestern University, Chicago, Illinois, USA.  
[simonwilliams@northwestern.edu](mailto:simonwilliams@northwestern.edu)

## Sexism: a revealing experiment

Few of the depressing statistics on women in science pinpoint hard evidence for bias against individuals (see, for example, *Nature* **495**, 22–24; 2013). So I conducted a small experiment of my own.

We know that successful grant applications are important drivers of promotion and tenure. In my first year as a researcher in 2005, I submitted 16 grant applications under my full name, which is not gender neutral. Just one received funding.

In my second year, I applied using only my initials and my last name: my success rate went up fivefold. This was only an  $n=1$  experiment, but I didn't

care to repeat it.

In my third year, the university adopted an electronic grant-application system, which, unbeknown to me, automatically entered my full first name. In this blinded study, my success rate went down fivefold again, coincident with changing back to a female name on the application.

The following year, I asked the university to modify my entry to use only initials for my first names, thereby frustrating the automated system. My success rate went back up fivefold.

I am the same applicant. The replicates are low, but the outcome apparently differed only when it was obvious to the reviewers that I was female.

**Tina M. Iverson** *Vanderbilt University Medical Center, Nashville, Tennessee, USA.*  
[tina.iverson@vanderbilt.edu](mailto:tina.iverson@vanderbilt.edu)

## Sexism: science biographer responds

In criticizing the stereotyping of female scientists, historian

Patricia Fara creates a new stereotype — for biographies of female scientists (*Nature* **495**, 43–44; 2013). She laments that they portray their subjects as weird, and protests against catchy titles, elements of cover design and the use of first names rather than surnames. As the author of one of the five books Fara criticizes, I would have welcomed a more substantive discussion. The challenges of writing scientific biography are more complex than she implies, whether the subject is female or male.

My book is, among other things, an exploration of the intertwined roles of truth and beauty in science, and of how a mathematician (Dorothy Wrinch) and a chemist (Linus Pauling) came to see them very differently. That difference is reflected in the title *I Died for Beauty: Dorothy Wrinch and the Cultures of Science*. I fail to see how, as Fara contends, this trivializes Wrinch's intellect.

**Marjorie Senechal Smith**  
*College, Northampton, Massachusetts, USA.*  
[senechal@smith.edu](mailto:senechal@smith.edu)



## NATURE'S READERS COMMENT ONLINE

A sample of responses from the debate on closing the gender gap in science (*Nature* **495**, issue 7439; 2013).

### James Dwyer says:

Simply quantifying inequities does not provide indisputable evidence of discrimination. A complete quantitative analysis would require that pay differences be normalized by some reliable measure of job performance — of value to the employer. Likewise, the number and value of research grants awarded should be normalized by not just the number of applications, but ideally by some independent measure of their quality.

### Julia Piaskowski says:

The '30-something' women you describe are astounding in their abilities to keep their careers spinning, but they set a very high bar. This is unrealistic for the average female scientist

and may have profound consequences for child-rearing.

### Helen Knob says:

As an associate professor at a high-ranking university and the mother of two young children, I see promising female graduate students and postdocs take off in flight from academia: they see what I am doing and think it is too hard. But running any enterprise with huge responsibilities as a full-time working mother is always going to be challenging, not least because 'full time' usually means almost every waking minute.

### Vivian Zapf says:

When my children arrived, I switched to part-time working. I thought that my career would

be permanently ruined. But by getting off useless committees, removing junk-work from my schedule, delegating and becoming more efficient, I found myself producing just as much science half-time as I had been full-time. And there is a reason why sabbaticals are built into the academic career: we all need time to step back and re-evaluate, and spending time with our children can provide just this opportunity.

### Anna Sutton says:

You know what's not helping? Quoting women saying "it's like not getting asked to dance". No wonder we're not asked to participate in technical and scientific activities if all we can do is focus on immature emotional reactions.

# Donald Arthur Glaser

(1926–2013)

Physicist and biotechnologist who invented the bubble chamber.

**D**onald Arthur Glaser was a great physicist and one of the first biotechnology entrepreneurs. He was also the least arrogant, most delightfully funny person you could hope to meet. He was aged just 25 when he designed and built the first bubble chamber, a device that allowed physicists to visualize with unprecedented precision the trajectory of elementary particles, and so measure their properties.

Glaser, who died on 28 February, was born in 1926, in Cleveland, Ohio. He received a bachelor's degree in physics and mathematics in 1946 from Cleveland's Case School of Applied Science (now Case Western Reserve University). As a child, he loved the arts — especially music — and while at Case he played the viola in the Cleveland Philharmonic Orchestra.

After his degree, Glaser started a PhD in physics at the California Institute of Technology (Caltech) in Pasadena. His interest in particle physics led him to work with Nobel laureate Carl Anderson, studying cosmic rays. To do this, he used a cloud chamber, an early particle detector that was essentially a sealed chamber filled with vapour. During these years, Glaser learned to design and build the equipment he needed for his experiments, a skill that served him well when he became an assistant professor at the University of Michigan in Ann Arbor in 1949.

It was at Michigan that Glaser built the first bubble chamber, an improvement on the cloud chamber. Whereas the cloud chamber was filled with vapour, the bubble chamber is filled with liquid. A piston produces a sudden decrease in pressure, forcing the liquid into a superheated state. Charged particles create an ionization track around which the liquid vaporizes, forming microscopic bubbles; the density of bubbles is proportional to the particle's energy loss. The device had a profound effect on particle physics, enabling the use of ever larger accelerators.

Glaser possessed a charming mix of curiosity and innocence. For a physicist considering how to visualize elementary particles, a natural step would have been to consult the relevant chapter in Enrico Fermi's 1937 book *Thermodynamics*, the reference text for particle physics. But instead Glaser worked through his own calculations from scratch. Had he read Fermi's book, he would have seen an equation that



implied that the bubble chamber could not work. Fermi had an iron-clad reputation for not making mistakes, but he had made one here, leaving the door open for Glaser's great invention.

In 1960, at the age of just 34, Glaser received the Nobel Prize in Physics for his work on the bubble chamber. Glaser's supervisor Anderson was the student of Robert Millikan, another physics Nobel prizewinner from Caltech. According to Glaser, his students used to look at one another and wonder who would be the fourth in this distinguished line. In 1993, the Nobel Prize in Chemistry did go to someone with whom Glaser had worked. That person was biochemist Kary Mullis, who received the prize for inventing the polymerase chain reaction (PCR), which can amplify one or a few pieces of DNA into thousands or millions of copies. Yet the connection between Glaser and Mullis had nothing to do with physics, but biotechnology.

Glaser became professor of physics at the University of California, Berkeley, in 1959. However, he grew tired of particle physics, which increasingly involved large teams of investigators using big instruments, and became professor of molecular biology

in 1964. His passion and skill at building equipment resulted in several new methods for automating large-scale hunts for pharmaceutically valuable mutants in organisms such as bacteria. He then did something that was unusual at the time. In 1971, with Ron Cape and Peter Farley, he co-founded Cetus, the first US biotechnology company, becoming chairman of its science advisory board. This was where Mullis discovered PCR, and other researchers synthesized interferon, a protein that helps to mediate the immune response. The company merged with Chiron in 1991 and was later purchased by Novartis.

As molecular biology became increasingly about biochemistry, Glaser again considered a career change. Partly because of his love of visual art, and because of problems he had encountered while training workers to recognize bubble-chamber tracks or assess the growth of cultures, he became fascinated by human vision. It was through this new interest that I met Glaser in 1981, and again in 1982 in Cambridge, Massachusetts, during his mini-sabbatical at the Rowland Institute — now part of Harvard University — where Edwin Land, the founder of Polaroid, was studying human perception of colour.

In the early 1980s, Glaser worked on visual psychophysics — a body of quantitative methods to measure perception — and on computational modelling of the visual system. From 1989, his position at Berkeley was professor of physics and neurobiology.

Don and I had stayed in close touch since 1981. I loved our discussions and the observations he made, which were simple, crisp and deep. It is both a privilege and a burden to win a Nobel prize at such a young age. And I always admired the grace with which Don carried his success in science and in industry. Somehow he retained the happy curiosity of a child: the world for him was a garden of wonders. He was able, in an apparently effortless way, to come up with refreshingly counterintuitive observations of things and of people. ■

**Tomaso Poggio** is professor in the Department of Brain and Cognitive Sciences at the McGovern Institute and at the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.  
e-mail: tp@csail.mit.edu

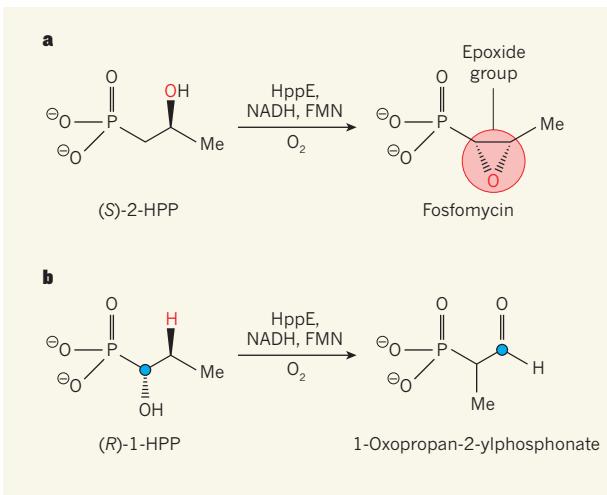
## Positive and radical

**Biosynthesis of the antibiotic fosfomycin ends with the formation of an ‘epoxide’ ring structure. The epoxide is suggested to form from a cationic intermediate, rather than from a free radical. SEE LETTER P.114**

SPENCER C. PECK  
& WILFRED A. VAN DER DONK

Most enzymes described in biochemistry textbooks catalyse reactions that involve the movement of two electrons. In such chemistry, a reagent (the nucleophile) with a reactive electron pair interacts with a partner (the electrophile) that is electron-deficient. Carbanions and carbocations are examples of carbon-based nucleophiles and electrophiles, respectively. In nature, reactions that cannot be easily accomplished by two-electron chemistry are often mediated by metalloenzymes that catalyse single-electron chemistry. In organic substrates, such reactions are typically thought to involve radicals (in which a carbon atom has one unpaired electron), but detecting these intermediates has been challenging because they are highly reactive. The non-haem iron-dependent enzymes form a large class of such metalloenzymes<sup>1</sup>. On page 114 of this issue, Chang *et al.*<sup>2</sup> provide data suggesting that at least one enzyme in this class uses both a carbon-centred radical and a carbocation as intermediates during the biosynthesis of fosfomycin, an antibiotic used in clinical settings and produced by several bacterial strains.

Fosfomycin contains a reactive epoxide group that is essential for its mode of action (Fig. 1a). The introduction of this oxygen-containing group during biosynthesis takes place through an unusual reaction catalysed by the enzyme (S)-2-hydroxypropylphosphonate epoxidase (HppE). Typically, the oxygen atom in naturally occurring epoxides derives from molecular oxygen ( $O_2$ ), and is added to a carbon–carbon double bond in the epoxide’s precursor molecule. However, *in vivo* experiments<sup>3</sup>, as well as *in vitro* incubation<sup>4</sup> of HppE with isotopically labelled (S)-2-hydroxypropylphosphonate, ((S)-2-HPP, the enzyme’s substrate), have demonstrated that fosfomycin’s epoxide oxygen derives from the alcohol group (OH) in 2-HPP. Since the first purification and



**Figure 1 | Reactions catalysed by the (S)-2-hydroxypropylphosphonate epoxidase (HppE) enzyme.** **a**, HppE catalyses the formation of the antibiotic fosfomycin from (S)-2-HPP by forming an epoxide group (circled). The oxygen atom in the epoxide comes from an alcohol group (OH) in the starting material, rather than from the oxygen ( $O_2$ ) used as a reagent. NADH and FMN are cofactors. **b**, Chang *et al.*<sup>2</sup> report that HppE converts a different substrate, (R)-1-HPP, into 1-oxopropan-2-ylphosphonate. The carbon atom shown with a blue dot ends up in the aldehyde group (CHO) of the product, and a hydrogen atom (known as the *pro-R* hydrogen; red) is lost.

definitive assignment of HppE as the epoxidase that carries out this reaction<sup>4</sup>, the enzyme has been extensively studied to uncover the catalytic mechanism involved.

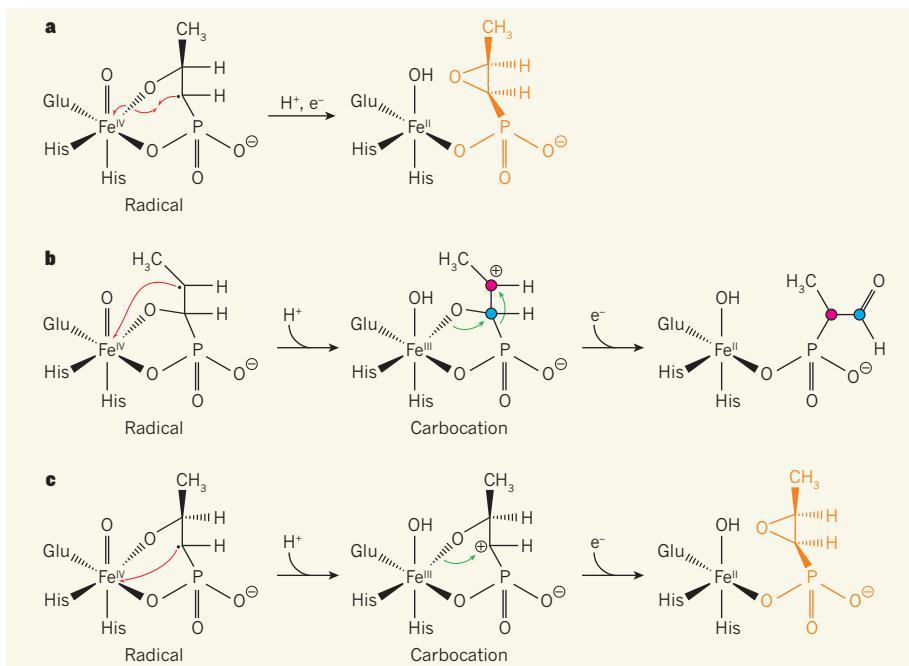
HppE needs oxygen and two electrons from a reducing agent to convert 2-HPP to fosfomycin. The coenzyme NADH, in the presence of the catalytic cofactor flavin mononucleotide, can serve as the two-electron donor to activate oxygen for the reaction *in vitro*<sup>4</sup>. Enzymology<sup>5</sup> and structural studies<sup>6</sup> have allowed the formulation of a putative mechanism in which a reactive iron species at the active site of HppE removes a hydrogen atom from the substrate to generate a carbon-centred radical. It was postulated that the epoxide forms when this radical reacts with 2-HPP’s deprotonated alcohol (the anion formed when the alcohol group has lost a proton,  $H^+$ ), which is bound to iron (Fig. 2a). The latter step could be described as an intramolecular version of the ‘oxygen rebound’ step that has been proposed for reactions catalysed by haem- and non-haem iron-dependent enzymes<sup>7</sup>.

Enzymatic reactions with substrates that involve one-electron chemistry are often well-controlled, and intermediates are difficult to detect because of their short-lived nature. Substrate analogues can be used to divert enzymes from their normal chemistry to an alternative outcome that can be diagnostic of a certain type of reactive intermediate. These analogues can therefore provide indirect evidence of the presence of an intermediate during the normal reaction.

In this spirit, Chang *et al.* discovered that incubation of HppE with the alternative substrate (R)-1-HPP generated 1-oxopropan-2-ylphosphonate (Fig. 1b). This product indicates that the phosphonate group ( $PO_3^{2-}$ ) in the substrate had moved to an adjacent carbon atom, in a process known as a 1,2-migration. Although this type of migration of a phosphorus-containing group has been observed before under ‘harsh’ conditions (such as with strongly basic reagents in organic solvents), its catalysis by an enzyme is unprecedented.

Chang and colleagues investigated the possible mechanism for the formation of the unexpected product by solving the crystal structure of HppE in complex with (R)-1-HPP. The structure showed that one of the hydrogens in the molecule (the *pro-R* hydrogen at the C2 position of (R)-1-HPP, Fig. 1b) is correctly oriented for removal by a reactive iron–oxygen species. To obtain further support for the idea that this hydrogen is removed, the authors incubated the enzyme with (R)-1-HPP in which the *pro-R* hydrogen was replaced with a deuterium atom. The product of the resulting reaction did not contain any deuterium, confirming the researchers’ hypothesis. The next question was: what type of reactive intermediate at C2 could go on to form the observed product?

To address this, the authors performed model studies. They used well-established chemical methods to generate either a radical or a carbocation at C2 of their substrate analogue in the absence of HppE, and



**Figure 2 | Proposed mechanisms of HppE-catalysed reactions.** **a**, In an earlier proposed mechanism for the HppE-catalysed synthesis of fosfomycin (orange), epoxide formation occurred through an intermediate radical. Fe, iron (Roman numerals indicate the oxidation state of the iron); e<sup>-</sup>, electron; Glu and His are the amino acids glutamate and histidine, respectively. The Fe=O species and the amino acids are part of the active site of HppE. The black dot is a single electron (the radical electron). Fish-hook arrows (red) indicate the movement of single electrons. **b**, Chang *et al.*<sup>2</sup> propose that in the final stages of the HppE-catalysed reaction of (R)-1-HPP, a one-electron transfer generates a carbocation, which rearranges in a 1,2-migration process to yield the product. Green arrows indicate movement of electron pairs. The phosphonate group ( $\text{PO}_3^{2-}$ ) moves from one carbon (blue) to the adjacent carbon (pink). **c**, The authors suggest that the final stages in the biosynthesis of fosfomycin might also involve a carbocation intermediate.

evaluated whether these intermediates gave rise to the same product as the enzyme-catalysed reaction. Their results show that, in organic solvents, a carbocation will indeed rearrange to generate the same product as the enzymatic reaction (Fig. 2b), whereas

the corresponding radical will not.

Chang and colleagues' data provide persuasive support for a carbocation intermediate in the enzymatic reaction of (R)-1-HPP. The carbocation might be formed by the transfer of an electron from an initially formed radical

to a highly oxidizing iron(IV) species (Fig. 2b). Further studies will be required to investigate whether such a carbocation intermediate is also formed during the reaction with 2-HPP (Fig. 2c). Moreover, Chang and co-workers' finding raises the question of whether other non-haem iron-dependent enzymes that use similar substrates but achieve different transformations<sup>8,9</sup> also generate carbocation intermediates. Regardless of the answer, the discovery of a 1,2-phosphono migration adds another reaction to the already impressive array of chemical transformations that are catalysed by non-haem iron-dependent enzymes in general<sup>1</sup>, and by HppE in particular. ■

**Spencer C. Peck and Wilfred A. van der Donk** are at the Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA, and in the Department of Chemistry and Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign.

e-mail: vddonk@uiuc.edu

1. Kovaleva, E. G. & Lipscomb, J. D. *Nature Chem. Biol.* **4**, 186–193 (2008).
2. Chang, W. *et al.* *Nature* **496**, 114–118 (2013).
3. Hammerschmidt, F. *J. Chem. Soc. Perkin Trans. 1*, 1993–1996 (1991).
4. Liu, P. *et al.* *J. Am. Chem. Soc.* **123**, 4619–4620 (2001).
5. Huang, H. *et al.* *J. Am. Chem. Soc.* **134**, 16171–16174 (2012).
6. Higgins, L. J., Yan, F., Liu, P., Liu, H. W. & Drennan, C. L. *Nature* **437**, 838–844 (2005).
7. Groves, J. T. *J. Inorg. Biochem.* **100**, 434–447 (2006).
8. Cicchillo, R. M. *et al.* *Nature* **459**, 871–874 (2009).
9. Metcalf, W. W. *et al.* *Science* **337**, 1104–1107 (2012).

off to become new North American real estate. In this issue, Sigloch and Mihalynuk<sup>2</sup> propose a new way of putting the pieces together, involving at least two additional oceanic plates. Their analysis is based on tomographic imaging of the deep mantle using earthquake waves, which reveals previously unseen features of the oceanic jigsaw.

Cordilleran rocks document a varied subduction history<sup>1,3</sup>. Plate convergence along the western North American margin probably started along an offshore subduction zone overlain by a chain of volcanoes arranged in an arc. The oceanic basin separating this volcanic arc from the mainland may have initially been closed by westward dipping subduction followed by eastward subduction below the continental margin. Several other intra-oceanic arcs and at least two larger ‘exotic’ terranes (blocks of crust deposited onto a plate), the Intermontane and Insular terranes, were subsequently accreted to North America’s west coast. From the mid-Cretaceous geological period about 90 million years

## EARTH SCIENCE

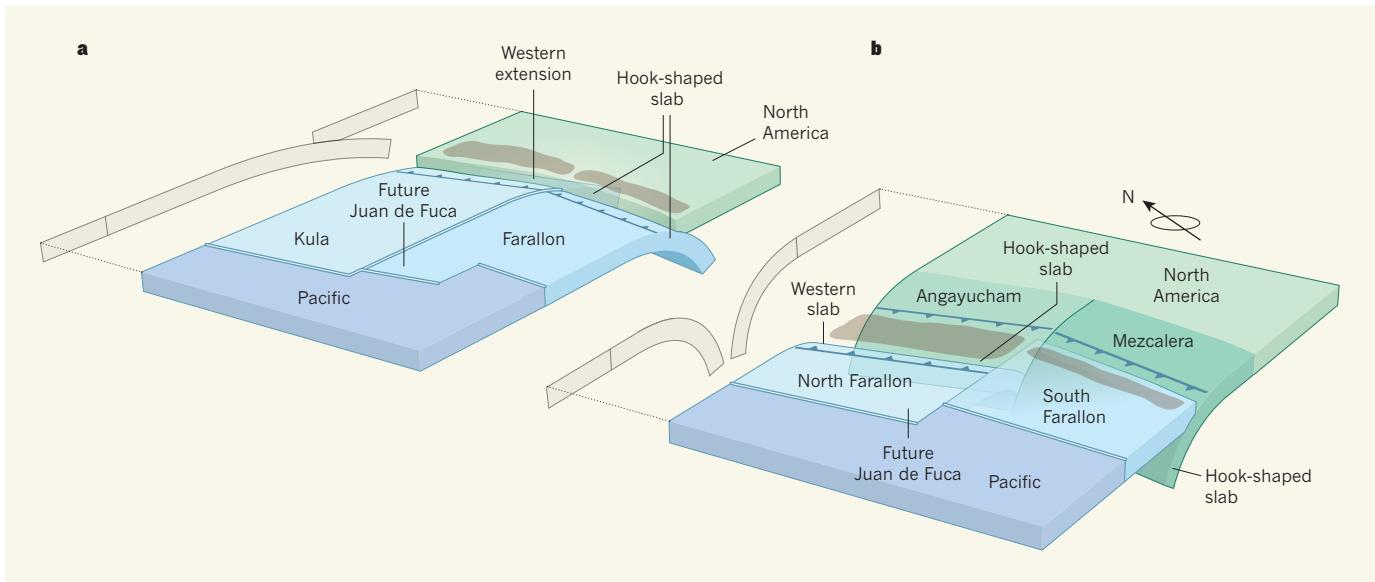
# Western North America's jigsaw

The tectonic history of western North America is a puzzle in which many of the pieces are mashed up or missing. Seismic images of the deep Earth reveal features that alter our reconstruction of this puzzle in space and time. SEE ARTICLE P.50

## SASKIA GOES

The western margin of the North American continent, comprising the Cordilleran mountain chain that runs from Alaska to Mexico, consists of scores of continental jigsaw pieces of different origins, sizes and ages. They came together as Cordilleran North America was assembled in a slow dance of tectonic plates over the past

350 million years<sup>1</sup>. It has been widely accepted that one large oceanic plate, the Farallon (and fragments thereof), acted as a conveyor belt, carrying all these pieces to North America’s western margin. When the plates converged, the dense oceanic plate sank into the mantle to form a slab below the lighter continent in a process called subduction. During subduction, volcanic island chains or small continental fragments on the oceanic plate were scraped



**Figure 1 | Two potential plate scenarios for North America's Cordilleran assembly.** **a**, The previously proposed plate configuration (corresponding to 80 million years ago), in which an eastward-subducting and fragmenting Kula/Farallon oceanic plate, bordered by the Pacific plate, rafted in all Cordilleran terranes (brown) and produced a hook-shaped mantle slab below the North American continent. A shallower section of the Kula plate formed a western extension of the hook-shaped slab. Only a remnant of the Farallon plate known as Juan de Fuca is currently subducting below North America. **b**, Sigloch and Mihalynuk's newly proposed scenario<sup>2</sup>

ago, a continental volcanic arc formed on the Cordilleran margin, sections of which continue to be active today. Sigloch and Mihalynuk link their tomographic imaging record of the deep mantle to the past 100–200 million years of Cordilleran assembly.

Reconstructing the assembly history of continental terranes is challenging because when plate slivers attach to a continent they are deformed and partly dismembered, and are further reshaped when additional pieces arrive. By contrast, the denser oceanic plates tend to remain largely intact during their tectonic cycle. Subducted oceanic-plate segments in the mantle therefore present a simpler image of plate history, although only mantle features with dimensions of 100 kilometres or larger can be resolved by earthquake tomographic imaging.

Sigloch and Mihalynuk's images reveal three major mantle slabs at a depth of between 1,000 and 1,800 km; the western slab stands alone and the two eastern ones are imaged as a continuous hook-shaped feature. This hook has previously been interpreted<sup>4,5</sup> as the result of the eastward subduction of a Farallon plate that broke at around 80 million years ago into a northern segment (the Kula plate) and a main Farallon core (Fig. 1). In previous tomographic images, the western slab was prominent only at depths above 1,500 km and was interpreted as a shallower section of a mildly dipping Kula slab<sup>5</sup>. An even shallower western slab was linked to a further (Juan de Fuca) fragment of the northern Farallon plate, which is currently

subducting at the Cascadia trench just offshore of the northwestern United States.

Surprisingly, the western feature in Sigloch and Mihalynuk's images seems to be a North Farallon slab that is continuous between the present-day Cascadia trench and a depth of 1,800 km. They therefore propose that the other deep slabs belong to non-Farallon oceanic plates, which they name the Angayucham and Mezcalera plates. These would have been attached to the North American continent and subducted westwards below a set of terranes that separated them from the Farallon plate (Fig. 1). Terrane docking ages, obtained from radiometric and fossil dating of the terrane rocks, require that the hook slabs predate the separation of the Kula and Farallon plates by about 60 million years, implying lower slab-sinking velocities and faster North American motion before 140 million years ago than if the imaged fragments were Kula–Farallon slabs.

The position of the features attributed to the South Farallon and Mezcalera slabs in the new tomographic images requires that the eastward subduction of the South Farallon ocean floor started at about the same time as the closure of the Mezcalera Ocean (the ocean underlain by the Mezcalera plate) was completed, depositing this Farallon slab on top of the Mezcalera one. Such subduction polarity switches have been proposed in Cordillera reconstructions before, as has the presence of the Mezcalera Ocean<sup>1</sup>. These subduction flips are usually attributed to the closure of small marginal basins, although a few studies have proposed

(corresponding to 140 million years ago), in which two westward-dipping oceanic plates, the Angayucham and Mezcalera, formed the hook-shaped slab, and a western slab formed by the North Farallon plate. Both scenarios are based on subducted slabs currently imaged in the deep mantle at a depth of 1,000–2,000 km. Divergent plate boundaries between the Pacific, Kula and Farallon plates are displayed as double lines; convergent boundaries are shown as lines with teeth in the direction of the downward-moving plate. The grey outlines show the side views of the left section of the plate configurations.

that most of 'Cordillera' assembled as a ribbon continent separated from North America by a large ocean<sup>6</sup>. None of these proposed oceans has previously been identified in the mantle record.

Sigloch and Mihalynuk have incorporated several new pieces into the North American jigsaw, but others remain to be placed. The authors have imaged several more westerly slabs that may require further intraoceanic subduction zones, for example. Some of these slabs have previously been interpreted as a record of further fragmentation of the Farallon plate below the southwestern United States<sup>7</sup>. Other studies have imaged slab material farther south and deeper in the mantle<sup>5</sup>, and these may be remnants of subduction before 150–200 million years ago. Could some of these correspond to a previously proposed<sup>1</sup> period of eastward subduction of the Mezcalera plate?

The new pieces remain to be connected into a self-consistent framework of plates with continuous and dynamically evolving boundaries. Subduction polarity flips are commonly associated with the arrival of buoyant terranes at the trench, for example at the closure of an ocean. What stopped earlier eastward subduction? Did the Mezcalera Ocean contain a ridge that clogged this subduction and left ocean-floor fragments stranded among the Intermontane terranes? Another intriguing question is whether a westward-subducting slab pulling on the North American plate might have aided the opening of the

Central Atlantic, which occurred in the same time frame as the proposed Angayucham-Mezcalera subduction. Many questions about the region's tectonic history remain, but this work highlights how tomographic images are providing increasingly well-defined pieces of the puzzle, complementing the oceanic and continental rock record. ■

**Saskia Goes** is in the Department of Earth Science and Engineering, Imperial College London, London SW7 2AZ, UK.

e-mail: [s.goes@imperial.ac.uk](mailto:s.goes@imperial.ac.uk)

- Dickinson, W. R. *Annu. Rev. Earth Planet. Sci.* **32**, 13–45 (2004).
- Sigloch, K. & Mihalynuk, M. G. *Nature* **496**, 50–56 (2013).
- Coney, P. J., Jones, D. L. & Monger, J. W. H. *Nature* **288**, 329–333 (1980).
- Bunge, H.-P. & Grand, S. P. *Nature* **405**, 337–340 (2000).
- Ren, Y., Stutzmann, E., van der Hilst, R. D. & Besse, J. *J. Geophys. Res.* **112**, B01302 (2007).
- Chamberlain, V. E. & Lambert, R. St J. *Nature* **314**, 707–713 (1985).
- van der Lee, S. & Nolet, G. *Nature* **386**, 266–269 (1997).

(essentially, micro-cracks following specific allowed paths within the crystals) and interfaces including low-angle grain boundaries and twins<sup>5</sup>. Researchers have imaged these dislocations using optical or electron microscopy, or X-ray-diffraction imaging techniques. Although they can image the extent of dislocations<sup>6,7</sup>, such techniques are incapable of imaging or producing meaningful measurements of the atomic structure of these defects.

The work by Chen *et al.* builds on a technique known as electron tomography. In this, a two-dimensional (2D) slice image of the sample under test is acquired using a fixed angle between the sample and the electron beam that passes through and undergoes diffraction. Thereafter, the sample is tilted at different angles with respect to this 'zero position' and a series of these 2D projections is recorded. The 2D images are then combined to produce a 3D image of the object. Until now, the ultimate spatial resolution of the 3D image has been limited by the steps in the sequential tilting of the sample, although this has been improved on by the development of what is known as equally sloped tomography (EST). In this approach, the tilt series of images is acquired by altering the tilt angle using an algorithm that ensures that the alteration occurs with equal slope increments<sup>8,9</sup>.

Another problem commonly encountered is that the acquired images are noisy (they possess a low signal-to-noise ratio), which greatly impedes the 3D rendering and impairs the spatial resolution. This signal-to-noise problem can be lessened by using mathematical operations known as Fourier transforms. These are used to relate the 3D *x*-*y*-*z* positional coordinates of the recorded image pixels to the 3D occurrence of their spatial-frequency components. Smooth variations in the pixel intensities across the image are termed low-frequency variations, whereas sharper, more regularly occurring ones are termed high-frequency variations. The Fourier-transform methods can extract this frequency information and filter those components that have a deleterious impact on the signal-to-noise ratio.

To overcome the signal-to-noise problem in their work, Chen *et al.* have subjected the EST images to Fourier transformation and used a 3D Fourier filtering method that identifies the Bragg peaks recorded on the EST images (Bragg peaks are electron-diffraction peaks that are the equivalent of the most commonly occurring spatial frequencies). This method also cleans up the data around the peaks and applies an inverse Fourier transform that converts the data back into the *x*-*y*-*z* space to produce spatially sharper 3D reconstructions.

The major contribution of Chen and colleagues' work is that it combines techniques that are already available, namely EST and 3D

## TECHNIQUES

# 3D imaging of crystal defects

A clever combination of existing techniques has produced three-dimensional atomic images of individual platinum nanoparticles, disclosing the atomic structure of crystal defects within them. SEE LETTER P.74

PATRICK J. McNALLY

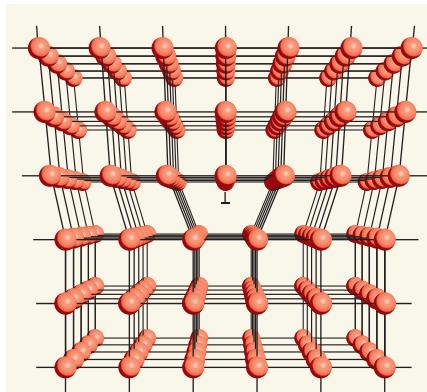
**C**rystals without defects are the 'Stepford wives' of materials science: all too perfectly, almost robotically, ordered, and relatively uninteresting. However, when one adds defects to these perfectly ordered solid-state arrangements of atoms or molecules, the materials acquire useful attributes, including the altered conducting or optical properties of semiconductor devices. But although well-tested theories exist to account for these defects at the atomic level, imaging them, and in particular doing so non-destructively, has proven a challenge. Chen *et al.*<sup>1</sup> (page 74 of this issue) have now made remarkable strides in this direction, producing three-dimensional (3D) images at atomic resolution of a number of structural defects inside a solid crystalline platinum nanoparticle\*.

Crystallography has a distinguished history of using macro-observational data — for example, measuring the interfacial angles in naturally occurring crystals such as rock salt or pyrite — to infer the existence at the atomic scale of 3D ordering in these solids. In the late eighteenth and early nineteenth centuries, French mineralogist René-Juste Haüy postulated that these naturally occurring crystals were composed of well-ordered assemblages of polyhedra, each of which was composed of elemental atoms<sup>2</sup>. Almost exactly 101 years ago, this was confirmed in the famous experiment<sup>3</sup> of Friedrich, Knipping and von Laue, in which a beam of Wilhelm Röntgen's newly discovered X-rays was directed at samples

\*This article and the paper under discussion<sup>1</sup> were published online on 27 March 2013.

of crystalline copper sulphate and zinc sulphide. The X-rays did not just pass through these crystals, but rather diffracted through them, producing ordered arrays of spots on a recording film. The location of these spots on the film corresponded to diffraction of the X-rays by individual, ordered atomic planes<sup>3,4</sup>, and confirmed the existence of atomic ordering in crystals.

Research in the intervening century has often tended to focus on the imperfections — the breakdown of local order — in these atomic arrays. Such imperfections can take the form of missing atoms (vacancies), atoms sitting in small gaps in these ordered arrays (interstitials) and many extended defects, of which the most common are dislocations



**Figure 1 | Edge dislocation in a cubic crystal.** This type of crystal defect is caused by the insertion of an extra atomic plane. Chen *et al.*<sup>1</sup> imaged edge dislocations at atomic resolution in a platinum nanoparticle that has a decahedral, rather than a cubic, atomic lattice.

Fourier filtering, to produce images of most of the platinum nanoparticle, which has a volume of approximately 850 cubic nanometres. The authors have proved the existence of atom-size steps at the boundaries between regions known as twins inside the nanoparticle (these twins comprise regions in which the atomic ordering within each twin is perfect but the twins themselves differ from each other in their local atomic arrangements). In addition, they have produced what I believe are the first images of the 3D atomic structure right at the core of two types of well-known dislocation — edge and screw dislocations — in the nanoparticle. Many of us remember materials-science lectures and textbooks in which we saw classical 3D block images of an edge or a screw dislocation (Fig. 1), but it is remarkable to see

this realized for individual atoms (see Figs 3 and 4 of the paper<sup>1</sup>).

Looking to the future, of particular interest for this 3D Fourier electron-tomography technique would be the development of a capability to map the exact location of every atom in a crystal. This is achievable to some extent in macromolecular or protein crystallography, although in these situations researchers have a priori knowledge of the molecular peptide sequence and stereochemistry, which helps greatly in refining the atomic positional modelling. The extension of 3D Fourier electron tomography to structures larger than nanoparticles should also be possible in principle, because the technique seems to be mainly constrained by electron scattering, and so imaging should depend chiefly on the electron-beam

energy and the composition of the material under test. ■

**Patrick J. McNally** is at the Nanomaterials Processing Laboratory, School of Electronic Engineering, Dublin City University, Dublin 9, Ireland.

e-mail: patrick.mcnamally@dcu.ie

- Chen, C.-C. et al. *Nature* **496**, 74–77 (2013).
- Whitlock, H. P. *Am. Miner.* **3**, 92–98 (1918).
- Friedrich, W., Knipping, P. & von Laue, M. *Sitzungsber. Bayer. Akad. Wiss. München* 303–322 (1912).
- Kubbinga, H. *Europhys. News* **43**(6), 20–23 (2012).
- Hull, D. & Bacon, D. *J. Introduction to Dislocations* 5th edn (Butterworth-Heinemann, 2011).
- Dash, W. C. *J. Appl. Phys.* **27**, 1193–1195 (1956).
- Allen, D. et al. *J. Appl. Cryst.* **44**, 526–531 (2011).
- Lee, E. et al. *J. Struct. Biol.* **164**, 221–227 (2008).
- Scott, M.C. et al. *Nature* **483**, 444–447 (2012).

gene expression to block cellular differentiation<sup>9–11</sup>. IDH mutations are therefore thought to promote cancer by modifying the epigenome to maintain cells in a stem-cell-like state.

In a series of compelling experiments, Losman *et al.* provide evidence that (R)-2HG transforms cells by inhibiting the activity of TET2. They find that blocking TET2 activity in cells recapitulated (R)-2HG-mediated growth-factor independence and impaired differentiation, reinforcing the suggestion that TET2 inhibition and the resulting epigenetic changes contribute to (R)-2HG-induced transformation.

However, the authors also identify one α-ketoglutarate-dependent dioxygenase whose activity must be maintained for (R)-2HG-mediated leukaemic transformation. This is EGLN, an enzyme that marks the transcription factor HIF-α for degradation (Fig. 1). This piece of the puzzle was elucidated when Losman *et al.* observed that the (S)-enantiomer of 2-hydroxyglutarate, (S)-2HG, does not promote leukaemia development, even though it inhibits TET2 and histone demethylases more potently than does (R)-2HG<sup>6,7</sup>. Therefore, it seemed probable that (R)-2HG promotes cellular transformation by modulating the activity of another dioxygenase.

It was recently demonstrated that (R)-2HG activates EGLN, whereas (S)-2HG inhibits it<sup>6</sup>. Consequently, Losman *et al.* reasoned that leukaemic transformation resulting from mutant IDH or (R)-2HG might require EGLN activity. Confirming this hypothesis, they observed that decreasing EGLN expression blocked transformation by mutant IDH or TET2 loss. Although the authors show that TET2 inhibition is sufficient for transformation and that EGLN activity is necessary for (R)-2HG-mediated transformation, whether TET2 inhibition is necessary for transformation by (R)-2HG remains to be tested.

The dependence of mutant IDH on EGLN activity during leukaemia development is surprising, as it contradicts the traditional view

## CANCER

# A metabolic metamorphosis

**Mutations in the enzyme isocitrate dehydrogenase lead to the accumulation of a metabolite that seems to promote cancer by influencing the epigenetic status of cells. But the effects are reversible, hinting at therapeutic targets.**

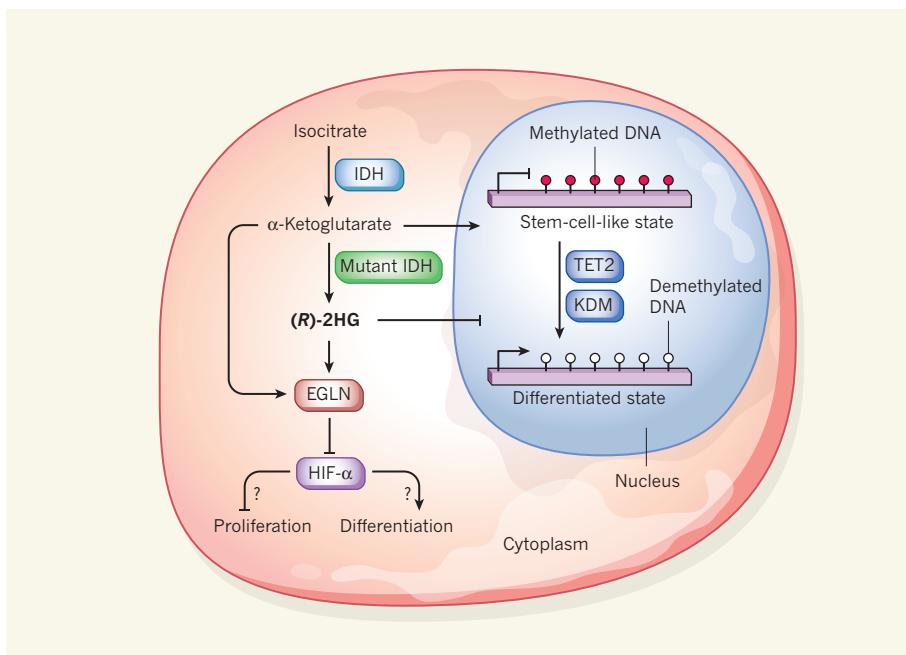
**ABIGAIL S. KRALL & HEATHER R. CHRISTOFK**

The discovery of cancer-associated mutations in metabolic enzymes has fuelled the idea that altered cellular metabolism may be involved in the development of cancer. For example, mutations in isocitrate dehydrogenases 1 and 2 are commonly detected in brain tumours and leukaemia. These mutations cause the enzymes to produce the (R)-enantiomer of a metabolite called 2-hydroxyglutarate, which can accumulate in tissues<sup>1,2</sup>. Writing in *Science*, Losman *et al.*<sup>3</sup> now show that the presence of this metabolite alone can promote cancer-like properties in cells. Notably, they also demonstrate that its cancer-promoting effects are reversible, suggesting that therapies targeting isocitrate-dehydrogenase mutants might improve patient prognoses.

Isocitrate dehydrogenases 1 and 2 (IDH1 and IDH2) function in the cytoplasm and in the citric-acid cycle, respectively, to convert isocitrate to α-ketoglutarate (Fig. 1). More than 75% of brain tumours and around 20% of acute myeloid leukaemias harbour mutations in IDH1 or IDH2<sup>4,5</sup> that confer on the enzymes the gain-of-function ability to convert α-ketoglutarate to (R)-2-hydroxyglutarate ((R)-2HG)<sup>1,2</sup>, which suggests that these mutations transform cells in ways that contribute to tumour development.

Losman and colleagues tested this idea by expressing a cancer-associated mutant version of IDH1 in immature red blood cells. They observed that the cells were able to grow in the absence of certain, usually essential, growth factors, and that their differentiation into mature cells was impaired — two hallmarks of leukaemic cells. Moreover, the addition of cell-membrane-permeable (R)-2HG to cells without an IDH mutation had the same effects, indicating that this metabolite is the transforming agent — a bona fide ‘oncometabolite’.

Dioxygenase enzymes that use α-ketoglutarate as a cofactor function in diverse cellular processes, such as adaptation to low oxygen levels or epigenetic modifications (alterations, such as methylation, to DNA or associated histone proteins that regulate gene expression without changing the DNA sequence). (R)-2HG, which structurally resembles α-ketoglutarate, promotes the activity of some of these enzymes by substituting for α-ketoglutarate<sup>6</sup>, and inhibits others by competing with α-ketoglutarate for active-site binding<sup>7</sup>. For example, (R)-2HG inhibits the dioxygenase TET2 (ref. 7), whose activity leads to reduced DNA methylation. It also inhibits some histone demethylase enzymes, such as KDM enzymes<sup>7,8</sup>. Inhibition of these epigenetic modifiers results in DNA and histone ‘hypermethylation signatures’ that can alter



**Figure 1 | An oncogenic metabolite.** Mutations in the metabolic enzyme isocitrate dehydrogenase (IDH), which converts isocitrate to  $\alpha$ -ketoglutarate can lead to accumulation of the normally irrelevant metabolite (R)-2HG. Losman *et al.*<sup>3</sup> show that the presence of (R)-2HG causes cells to become independent of growth factors and to lose the ability to progress from a stem-cell-like progenitor to a fully differentiated state — two hallmarks of leukaemic cells. The authors' data suggest that this occurs because the metabolite impairs the activity of  $\alpha$ -ketoglutarate-dependent enzymes such as TET2 and KDM, which promote differentiation by removing methyl groups from DNA and DNA-associated histone proteins to alter gene expression. The authors also show that (R)-2HG-mediated cellular changes depend on the activity of EGLN, another enzyme that uses  $\alpha$ -ketoglutarate as a cofactor. EGLN marks the transcription factor HIF- $\alpha$  for degradation. However, HIF- $\alpha$  is typically associated with cancer promotion, so the role of EGLN in (R)-2HG-mediated cancer promotion remains unclear.

of HIF-1 $\alpha$  as a cancer-promoting protein. However, in support of Losman and colleagues' findings, previous studies have shown that HIF-1 $\alpha$  can inhibit leukaemic-cell proliferation and induce the differentiation of myeloid cells<sup>12</sup>. This occurs in the absence of its typical transcription-factor partner HIF-1 $\beta$ , and seems to depend instead on CEBP $\alpha$ , a transcription factor that is crucial for blood-cell differentiation<sup>13</sup>. Further studies are needed to determine whether HIF- $\alpha$  loss is a major factor in (R)-2HG-mediated transformation.

Importantly, Losman *et al.* also demonstrate that the oncogenic effects of (R)-2HG are reversible — its removal restored the growth-factor dependence and differentiation capacity of cells even after long-term exposure to the metabolite. This reversibility is a key observation because mutant IDH may transform cells by altering the epigenome. The stable, inheritable nature of epigenetic modifications had raised concerns that targeting mutant IDH would not be therapeutically effective. But if (R)-2HG removal also results in reversion of hypermethylation signatures, this would indicate that the epigenetic modifications relevant to mutant IDH-mediated transformation are more dynamic than previously thought. Alternatively, however, the features of transformed

cells may arise from atypical functions of epigenetic modifiers such as TET2, or solely from the activities of other enzymes affected by (R)-2HG, such as EGLN.

Mounting evidence suggests that metabolism and cellular epigenetic states are interconnected: because many epigenetic modifiers depend on metabolites as cofactors or substrates, the cell's metabolic state can influence its epigenetic state. Moreover, epigenetic modifications regulate many cellular processes central to cancer development. The study by Losman *et al.* bridges the fields of cancer, metabolism and epigenetics by showing that a metabolite that results from a genetic abnormality can promote cancer, possibly by changing the epigenetic landscape of the cell to alter gene expression. Continued study of the dynamic interplay between metabolism and epigenetics in cancer should lead to improved understanding of disease development and better therapeutic options. ■

**Abigail S. Krall and Heather R. Christofk**  
are in the Department of Molecular  
and Medical Pharmacology, University  
of California, Los Angeles,  
Los Angeles, California 90095, USA.  
e-mails: [akrall@ucla.edu](mailto:akrall@ucla.edu);  
[hchristofk@mednet.ucla.edu](mailto:hchristofk@mednet.ucla.edu)

1. Dang, L. et al. *Nature* **462**, 739–744 (2009).
2. Ward, P. S. et al. *Cancer Cell* **17**, 225–234 (2010).
3. Losman, J. A. et al. *Science* <http://dx.doi.org/10.1126/science.1231677> (2013).
4. Parsons, D. W. et al. *Science* **321**, 1807–1812 (2008).
5. Mardis, E. R. et al. *N. Engl. J. Med.* **361**, 1058–1066 (2009).

6. Koivunen, P. et al. *Nature* **483**, 484–488 (2012).
7. Xu, W. et al. *Cancer Cell* **19**, 17–30 (2011).
8. Lu, C. et al. *Nature* **483**, 474–478 (2012).
9. Figueroa, M. E. et al. *Cancer Cell* **18**, 553–567 (2010).
10. Turcan, S. et al. *Nature* **483**, 479–483 (2012).
11. Moran-Crusio, K. et al. *Cancer Cell* **20**, 11–24 (2011).
12. Song, L.-P. et al. *Oncogene* **27**, 519–527 (2008).
13. Friedman, A. D. *Oncogene* **21**, 3377–3390 (2002).

H.R.C. declares competing financial interests: see go.nature.com/8p7lek for details.

## MATERIALS SCIENCE

# The same, but better

**Moving beyond mimicry, biologically inspired artificial materials can be simpler in design yet more powerful in function than their natural analogues. A tropical fruit seed serves as a guide to making new photonic elements.**

TERI W. ODOM

**N**ature's bounty has provided the inspiration for a wide range of man-made photonic materials. For example, opals have given rise to artificial crystals made of colloidal building blocks<sup>1</sup>, and the starfish-related brittlestar has shown how microlenses can act like a compound eye<sup>2</sup>. Now, as Kolle *et al.* report<sup>3</sup> in *Advanced Materials*, tropical plant seeds can be added to the list, and with them, a new biologically inspired photonic element — the soft optical fibre. The seeds of the rainforest plant *Margaritaria nobilis* look blue-green because of the way in which light interacts with the cells in the plant's seed coat, which are elongated, with a cross-section rather like a Swiss roll in structure (Fig. 1a,b). This hierarchical architecture, in which the overall curvature of the cell is microscale and the concentric features within the cell are nanoscale, provided the guiding principles for the authors' design of a multilayered photonic fibre (Fig. 1c).

What is significant about Kolle and colleagues' work is that they did not simply mimic the cellular structure of the coat, and hence obtain similar blue-green colours. Instead, they created a structure that had a simpler geometry which could not only produce blue and green colours, but also generate tunable colours over the visible spectrum. The key to this greater flexibility resides in the use of soft materials and in the fabrication method.

The colours of the seed coat are determined by the thickness and refractive index of the two alternating materials of the Swiss-roll structure, whereas the brightness is controlled primarily by the number of bilayers in the roll. When light shines on a multilayered stack of two alternating materials, the reflected colour — whose frequency falls within a range of 'forbidden' frequencies called the photonic bandgap — is preserved only at angles perpendicular to the surface. But because the cells in the seed coat have microscale curvature in

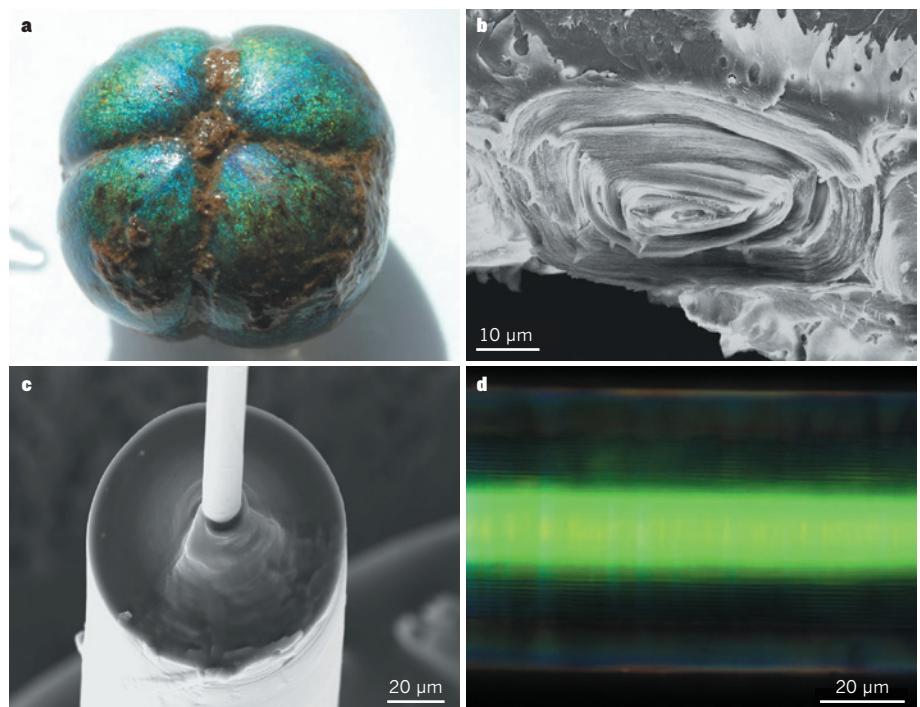
addition to the concentric nanoscale multilayers, and the fruit itself has macroscale curvature, the blue-green colours can be observed over a wider range of angles.

To create a bilayer with different refractive indices, Kolle *et al.* used two polymers, polydimethylsiloxane (PDMS) and polystyrene-polyisoprene triblock copolymer (PSPI). The thicknesses of the two polymer layers determined the central frequency of the photonic bandgap. For example, a PDMS film 240 nanometres thick and a PSPI 100 nm thick that were rolled around a 14-micrometre-diameter glass rod nearly 80 times resulted in a green fibre (Fig. 1d). With its circular cross-section, this soft fibre could reflect green colour without

needing the structural complexities of the seed coat, such as the cells' elliptical cross-section and the fine structure within their nanoscale layers.

Kolle *et al.* took advantage of the elastic properties of the polymer layers to tune the optical properties of the soft fibre. After the glass rod was removed, the now-hollow fibre could be mechanically deformed. Stretching along the axis of the fibre was accompanied by compression in the radial direction according to Poisson's ratio for elastic materials under strain. Therefore, the authors could tune the fibre colour from the starting colour to shorter wavelengths simply by stretching. Tuning the fibre colour to longer wavelengths could potentially be achieved by inducing swelling, and thereby increasing the thickness of the polymer layers, by exposure to organic solvents. This type of flexibility is not possible in the tropical fruit seed coat or in other natural structures made from hard materials.

Despite the potential of these soft photonic fibres, three main challenges remain: colour uniformity, colour purity and scalability. For the first problem, when the fibre is stretched along its axis, the strain is not uniformly



**Figure 1 | Fruit-inspired optical fibres.** **a**, The fruit of the tropical plant *Margaritaria nobilis* without its capsule (about 10 mm in diameter). **b**, Cross-section of the fruit seed coat shows a concentric, multilayered cellular structure. **c**, Cross-section of a soft multilayered optical fibre around a glass core, fabricated by Kolle and colleagues<sup>3</sup>. **d**, Reflected green colour from the soft optical fibre in **c**. REF. 3

distributed along the length of the fibre and is most different where the fibres are clamped for stretching. Because the strain is non-uniform, the reflected colour will be also. In addition, if there is any slippage between the bilayers as the fibre is strained, there will be non-uniform colour in random areas. For the second issue, having a circular rather than an elliptical cross-section results in a larger gradient in colour variation. This could be overcome by starting the rolling process with a rod that has an elliptical or rectangular cross-section. For the third concern, the rolling of bilayers around a glass-rod template is cumbersome, and the rod must later be removed to obtain tunable colour. This might be addressed by pre-straining the bottom layer in the bilayer such that the released strain would result in self-rolling with the appropriate inner diameter. ■

Overall, however, the versatility of Kolle and colleagues' fabrication approach bodes well for a range of future applications. For example, if insulator–metal bilayers can be rolled, and if the free-standing metal layer were to contain nanoscale patterns<sup>4</sup>, then fibre metamaterials could be formed. Such fibres could increase speeds for telecommunications and sensing. In addition, the authors'

technique could be used to make multilayer claddings out of many more combinations of materials than those limited by standard fibre drawing, which can be done only at high temperatures<sup>5</sup>. The room-temperature processing of the soft fibres is a distinct advantage. Finally, these flexible colour fibres could be woven into garments used for camouflage by subjecting them to external stimuli such as strain or volatile solvents. The re-engineering of nature's resources for photonic materials with enhanced capabilities has never looked more fruitful. ■

**Teri W. Odom** is in the Departments of Chemistry and of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA.  
e-mail: todom@northwestern.edu

1. Velev, O. D., Jede, T. A., Lobo, R. F. & Lenhoff, A. M. *Nature* **389**, 447–448 (1997).
2. Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L. & Helder, G. *Nature* **412**, 819–822 (2001).
3. Kolle, M. *et al.* *Adv. Mater.* <http://dx.doi.org/10.1002/adma.201203529> (2013).
4. Henzie, J., Lee, M. H. & Odom, T. W. *Nature Nanotechnol.* **2**, 549–554 (2007).
5. Varshneya, A. K. *Fundamentals of Inorganic Glasses* (Academic, 1993).

#### BIOCHEMISTRY

## Sirtuin on a high-fat diet

**It emerges that the sirtuin enzyme SIRT6 preferentially removes long-chain fatty-acyl, rather than acetyl, protein modifications. This activity regulates secretion of the inflammation-associated protein TNF- $\alpha$ .** SEE LETTER P.110

POONAM BHEDA & CYNTHIA WOLBERGER

**S**irtuins regulate a broad range of processes, including transcription, metabolism, fat mobilization, neurodegeneration and ageing. The various functions of these proteins have been largely ascribed to their ability to catalyse the removal of acetyl groups from the lysine amino-acid residues of other proteins through their deacetylase activity. Yet the exact biological action of sirtuins remains unclear. For instance, one sirtuin, SIRT6, which has been implicated in genome stability, inflammation, cancer-cell metabolism and even lifespan, is a very weak deacetylase<sup>1</sup>. On page 110 of this issue, Jiang *et al.*<sup>2</sup> report the surprising discovery that SIRT6 robustly removes a myristoyl group — a long-chain fatty-acyl group — from lysine residues, and that this biochemical activity enables the enzyme to regulate the secretion of TNF- $\alpha$ , a cytokine protein released from cells during inflammation.

Proteins undergo a diverse array of chemical modifications that modulate their activity. The

enzymes that add and remove these modifications are therefore key decision-makers in signalling cascades. The lysine side chains of proteins can be modified by attachment of a small acyl group called acetyl, which is one of the most common regulatory modifications and is best known for its role in controlling transcription. Other, larger acyl modifications of lysine residues have been detected, although their biological roles are largely unknown.

Most sirtuins deacetylate lysine side chains in a reaction that consumes the cofactor nicotinamide adenine dinucleotide (NAD<sup>+</sup>), releasing nicotinamide and transferring the acetyl group from the lysine to the remaining ADP-ribose moiety of NAD<sup>+</sup> to form O-acetyl ADP-ribose (Fig. 1). Most of the seven human sirtuins (SIRT1–7) exhibit this typical activity, although some, including SIRT6, have either weak or no deacetylase activity. For instance, SIRT5 preferentially binds to and removes succinyl and malonyl modifications from lysine<sup>3</sup>. These acyl groups are larger than acetyls and, unlike them, are negatively charged, but they are linked to lysine by the



## 50 Years Ago

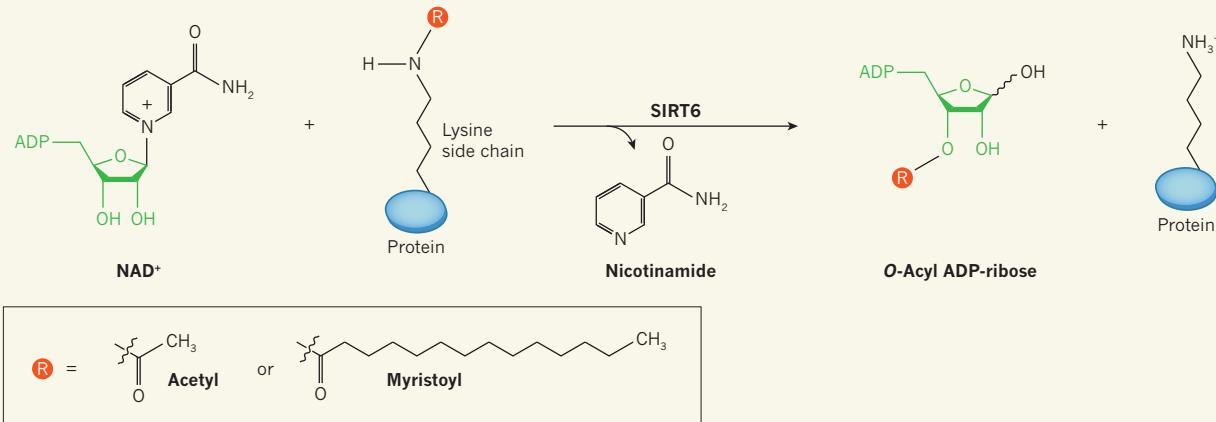
It is surprising to find that despite the widespread attention directed to the United Nations conference last February on the application of science and technology for the benefit of the less-developed areas, the admirable issue "World of Opportunity" of the *New Scientist* of February 14 still appears to remain the only one in which that conference has been at all adequately handled by British periodicals. This comparative neglect, in spite of copious literature distributed by the Department of Scientific and Industrial Research, in no way detracts from the excellence of the issue of the *New Scientist* which is as welcome for its enterprise as for its intrinsic and timely merit.

From *Nature* 6 April 1963

## 100 Years Ago

Disastrous floods have followed the severe windstorms in the United States on March 23. The areas chiefly affected are the middle western States ... Immense tracts of country have been submerged, and many large towns have become flooded. Much of the ground is below the flood level of the rivers, and in parts the embankments have given way, whilst many tributary rivers have overflowed their banks. Dayton, Indianapolis, Columbus, and numerous other smaller towns have suffered immensely during the last week of March and much loss of life has occurred. Hundreds of houses have been washed away, and immense suffering has been caused ... The rivers are said to be still rising in many places, and the full result of the disaster will depend largely upon the weather for the next week or two.

From *Nature* 3 April 1913



**Figure 1 | Protein deacylation.** Sirtuins are known to act as deacetylase enzymes. They remove the acyl group (R) acetyl from the lysine side chain of substrate proteins in a reaction involving the  $\text{NAD}^+$  cofactor and transfer it to the remaining ADP-ribose moiety of  $\text{NAD}^+$  (green) to form

O-acetyl ADP-ribose. (Nicotinamide is a by-product of the reaction.) Jiang *et al.*<sup>2</sup> report that SIRT6 preferentially removes a different, long-chain acyl group (myristoyl) from proteins — a finding that, together with previous data, call for a reclassification of sirtuins as deacylases.

same type of chemical bond and are removed in the same  $\text{NAD}^+$ -dependent enzymatic reaction as that catalysed by other sirtuins.

The SIRT5 precedent prompted Jiang *et al.* to investigate whether the apparently weak deacetylase activity of SIRT6 also reflects a preference for other acyl-lysine substrates. Their search began *in vitro* with the use of chemically synthesized peptides bearing lysines modified with various acyl groups that are known to occur in cells. The result was clear: SIRT6 was far more active in removing the long-chain fatty-acyl myristoyl and palmitoyl groups than small acyl modifications, including acetyl groups.

The authors' crystal structure of SIRT6 bound to a myristoylated peptide and ADP-ribose shows that SIRT6 contains an expanded hydrophobic channel in which it can accommodate the 14-carbon myristoyl chain. Together with previous structural studies<sup>3–5</sup> showing how different sirtuins can accommodate various acyl modifications, including succinyl, malonyl and propionyl lysines, it now seems clear that structural features in the sirtuins' active sites govern the preference of each one for removing a particular type of acyl modification.

*In vivo*, SIRT6 is known to regulate the levels of TNF- $\alpha$ , which spans the cell membrane and is cleaved by a membrane-associated protease enzyme, resulting in the secretion of this cytokine's extracellular domain. The cytoplasmic domain of TNF- $\alpha$  contains two myristoylated lysines. This immediately raises two questions: does SIRT6 remove these myristoyl modifications and, if so, is this enzymatic activity somehow connected to TNF- $\alpha$  regulation?

Indeed, Jiang *et al.* found that the myristoylation level of TNF- $\alpha$  in cultured cells depended on the enzymatic activity of SIRT6. Secretion of TNF- $\alpha$  also required SIRT6,

indicating that removal of its myristoyl groups is a key step in this process. It will be interesting to see how myristylation regulates secretion, and whether removal of the myristoyl groups causes a conformational rearrangement in TNF- $\alpha$  that allows its cleavage by the membrane-associated protease.

Jiang and colleagues' findings set the stage for several new directions in which to investigate the role of fatty-acyl modifications and their regulation by sirtuins. The paper should be the final impetus for reclassifying sirtuins as lysine deacylases<sup>5</sup>, and not simply deacetylases, to reflect the more general nature of their enzymatic activity.

Importantly, it now seems that sirtuins differ from one another in the type of acyl modification they preferentially remove from substrates, although current findings do not rule out the possibility that a given sirtuin can remove several types of acyl modification *in vivo*. This is perhaps the case for SIRT6. Although the enzyme preferentially removes long-chain fatty-acyl modifications, its deacetylase activity has been implicated<sup>6</sup> in modification of the DNA-associated histone H3 protein. It could be that, *in vivo*, the weak deacetylase activity of SIRT6 is restricted to specific substrates (as has been shown<sup>6</sup>), to precise subcellular localizations or to particular signalling pathways. The fact that SIRT6 can remove a broader spectrum of acyl modifications should make it possible to tease out the relative contributions of these biochemical activities to this enzyme's function.

It has long been a mystery why sirtuins consume the energetically costly  $\text{NAD}^+$  cofactor, rather than — like other classes of deacetylase — using simple hydrolysis to deacetylate substrates. Explanations invoked include the regulatory implications of coupling sirtuin activity to the cell's metabolic state, in which

$\text{NAD}^+$  is involved, or the possible signalling role of the O-acetyl ADP-ribose product (Fig. 1). Another explanation may be the need to generate a nucleotide carrier for the departing acyl group. On transferring from the lysine substrate to ADP-ribose to form O-myristoyl ADP-ribose, the hydrophobic myristate is released in a more soluble conjugated form. The O-myristoyl ADP-ribose conjugation also prevents myristate from becoming elongated by further enzymatic reaction to form another fatty acid, palmitate, thereby preserving the available pool of myristate.

O-Acetyl ADP-ribose is broken down by several enzymes<sup>7</sup>, and it remains to be seen which of them regulate the levels of other O-acetyl ADP-ribose products or transfer them to other carriers, and whether these activities affect or even drive upstream pathways involved in generating these metabolites. For now, the discovery of SIRT6-mediated demyristoylation opens an exciting chapter in the story of the seven human sirtuins and their biological activity. ■

**Poonam Bheda** is at the Institut Génétique et de Biologie Moléculaire et Cellulaire, Illkirch 67404, France. **Cynthia Wolberger** is in the Department of Biophysics and Biophysical Chemistry and the Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205-2185, USA.  
e-mail: cwolberg@jhmi.edu

- Lerner, B. & Cohen, H. Y. *EMBO J.* **32**, 7–8 (2013).
- Jiang, H. *et al.* *Nature* **496**, 110–113 (2013).
- Du, J. *et al.* *Science* **334**, 806–809 (2011).
- Bheda, P., Wang, J. T., Escalante-Semerena, J. C. & Wolberger, C. *Protein Sci.* **20**, 131–139 (2011).
- Zhu, A. Y. *et al.* *ACS Chem. Biol.* **7**, 155–159 (2012).
- Michishita, E. *et al.* *Nature* **452**, 492–496 (2008).
- Tong, L. & Denu, J. M. *Biochim. Biophys. Acta* **1804**, 1617–1625 (2010).

# Patterns and mechanisms of early Pliocene warmth

A. V. Fedorov<sup>1,\*</sup>, C. M. Brierley<sup>1,2,\*</sup>, K. T. Lawrence<sup>3\*</sup>, Z. Liu<sup>4</sup>, P. S. Dekens<sup>5</sup> & A. C. Ravelo<sup>6</sup>

About five to four million years ago, in the early Pliocene epoch, Earth had a warm, temperate climate. The gradual cooling that followed led to the establishment of modern temperature patterns, possibly in response to a decrease in atmospheric CO<sub>2</sub> concentration, of the order of 100 parts per million, towards preindustrial values. Here we synthesize the available geochemical proxy records of sea surface temperature and show that, compared with that of today, the early Pliocene climate had substantially lower meridional and zonal temperature gradients but similar maximum ocean temperatures. Using an Earth system model, we show that none of the mechanisms currently proposed to explain Pliocene warmth can simultaneously reproduce all three crucial features. We suggest that a combination of several dynamical feedbacks underestimated in the models at present, such as those related to ocean mixing and cloud albedo, may have been responsible for these climate conditions.

Earth's climate evolution over the last five million years (Myr), since the early Pliocene epoch, has been meticulously studied by scientists (see, for example, refs 1–3). For years, the focus of attention was on the origin<sup>4–6</sup> of glacial cycles, that is, the coming and going of continental ice sheets evident in the δ<sup>18</sup>O record reflecting global ice volume and deep-ocean temperature (Fig. 1a). However, a wealth of new data now show that the gradual onset (around 3 Myr ago) and further amplification of Northern Hemisphere glaciation (Fig. 1a, b) was only one facet of this climate change and was a consequence of the global cooling rather than its initial cause<sup>7,8</sup>.

The early Pliocene itself, the warm interval that preceded the glaciation, has attracted a lot of attention as a possible analogue for future climate conditions. Despite relatively small differences in climate control factors, including CO<sub>2</sub> concentrations (Fig. 1c), between the early Pliocene and the present, the former was markedly different<sup>9</sup>. Palaeorecords indicate vast changes in climate patterns since 5–4 Myr ago, including a contraction of the tropical belt and oceanic warm pool<sup>10</sup>, emergence of strong temperature gradients along the Equator<sup>11</sup>, cooling of coastal upwelling zones in the subtropics<sup>12</sup>, the shoaling of the ocean thermocline<sup>11,13</sup> and cooling of the high-latitude and deep ocean<sup>14,15</sup>. Together these observations imply a large structural change in climate, with major global and regional implications.

By a structural change, we mean a transition from a climate with almost no zonal sea surface temperature (SST) gradients (~1 °C or less) and a weak meridional SST gradient to a modern climate with much more pronounced spatial temperature contrasts. This structural climate change was paralleled by relatively small changes in atmospheric CO<sub>2</sub> concentration (Fig. 1c), on the order of 50–100 p.p.m. Thus, a key step towards understanding early Pliocene climate, and possibly other warm climates throughout Earth history, is to explain what could cause such changes.

The objectives of this article are twofold. First we review recent geochemically derived palaeorecords emphasizing the three criteria that any hypothesis must satisfy to account for the climate of the early Pliocene (reduced meridional temperature gradient, weak zonal gradients and SST stability). Then we evaluate various explanations for

early Pliocene climate with respect to these criteria, using a single modelling framework.

## Observations

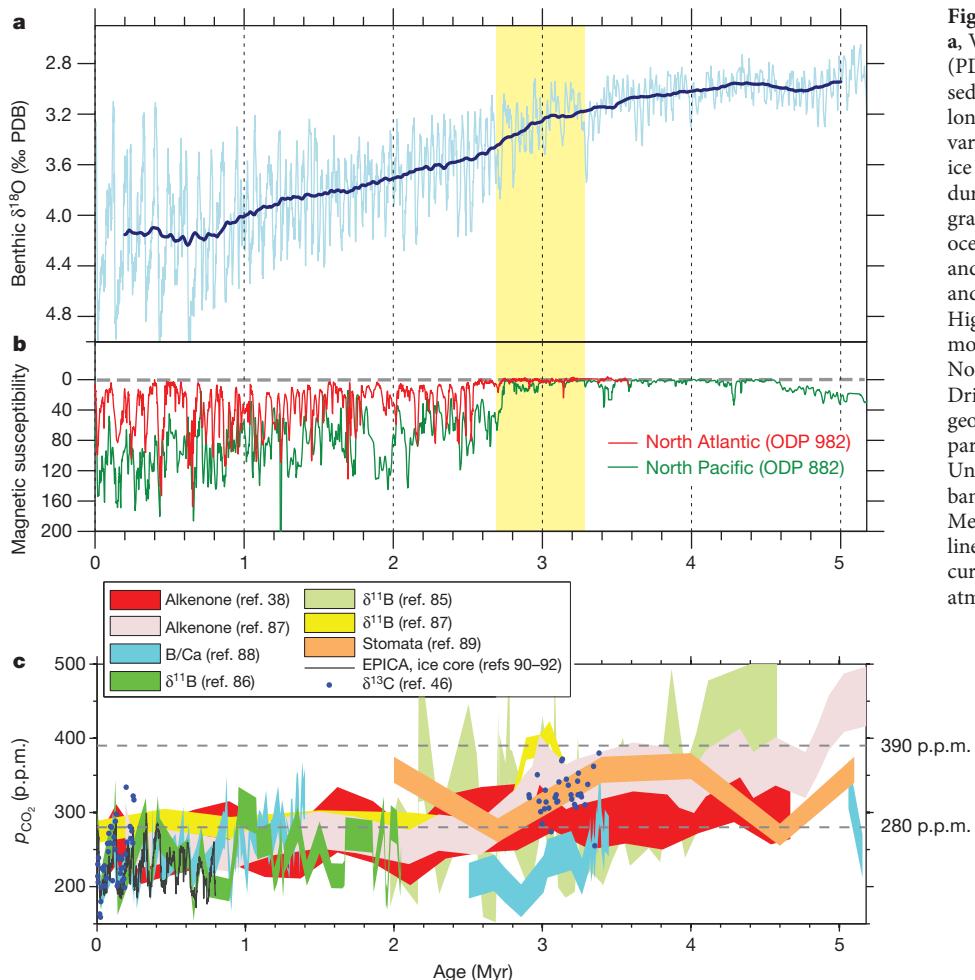
Substantial effort has been put into deriving palaeoclimate records for the Pliocene because it is recognized as the most recent example of prolonged global warmth in the geological past<sup>2,16</sup>. As part of the PRISM project, this effort produced a series of global temperature reconstructions<sup>17,18</sup> for a ‘time slab’ of the mid Pliocene (3–3.3 Myr ago). These data sets and associated modelling studies enable a coordinated comparison between the Pliocene and the present, and dramatically increased our understanding of Pliocene climate.

In the present study, we compile available SST proxy records, most of them continuous, to describe the long-term climate trends from the warm early Pliocene to the present (Figs 2 and 3, and Methods). As several of these records indicate, the period from 4.4 to 4 Myr ago was probably the warmest interval within this timeframe—the Pliocene climatic optimum (Fig. 2b, c). The subsequent cooling became evident roughly at the same time in both hemispheres (Fig. 2c, d) and involved regions ranging from low-latitude upwelling zones to mid and high latitudes. The expansion of the Northern Hemisphere ice sheets around 2.7 Myr ago, evidenced by an increase in the magnetic susceptibility of sediments affected by ice-raftered debris (Fig. 1b), introduced a strong interhemispheric asymmetry in the climate evolution over land<sup>19</sup> but to a lesser degree over the ocean (as shown by the similarity of trends in Fig. 3c, d).

Within the global cooling pattern, local trends featured significant spatial and temporal variations not necessarily tracking the growth of continental ice sheets inferred from δ<sup>18</sup>O data (Fig. 1a). Regionally, temperature changes were as high as 11 °C over the ocean<sup>20</sup> (Fig. 3c) and 19 °C over the land<sup>19</sup>. Below, we focus on the critical features that characterize the evolution of climate structure since the early Pliocene. As is traditionally done by the palaeoclimatology community, we compare the mean trends incorporating both glacial and interglacial intervals of the time series, even though some of our statements contrast the late Quaternary interglacials and Pliocene interglacials. A

<sup>1</sup>Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA. <sup>2</sup>Department of Geography, University College London, London WC1E 6BT, UK. <sup>3</sup>Department of Geology and Environmental Geosciences, Lafayette College, Easton, Pennsylvania 18042, USA. <sup>4</sup>Department of Earth Sciences, The University of Hong Kong, Hong Kong. <sup>5</sup>Department of Geosciences, San Francisco State University, San Francisco, California 94132, USA. <sup>6</sup>Department of Ocean Sciences, University of California, Santa Cruz, California 95064, USA.

\*These authors contributed equally to this work.



**Figure 1 | Climate evolution over the past 5 Myr.** **a**, Variations in benthic  $\delta^{18}\text{O}$  (PeeDee Belemnite (PDB) standard; Methods) from deep-ocean sediments, showing glacial cycles (light blue) and a long-term trend<sup>50</sup> (dark blue).  $\delta^{18}\text{O}$  records past variations in deep-ocean temperature and global ice volume. The shading highlights the period during which Northern Hemisphere glaciation gradually begins<sup>5</sup>. **b**, Magnetic susceptibility of ocean sediments in the northern Pacific<sup>51</sup> (green) and Atlantic<sup>52</sup> (red). The units are nondimensional and the data are presented as originally published. Higher susceptibility indicates the presence of more ice-rafter debris, corresponding to larger Northern Hemisphere ice sheets. ODP, Ocean Drilling Program. **c**, Proxy  $p\text{CO}_2$  data from marine geochemical studies and ice-core inclusions ( $p\text{CO}_2$ , partial pressure of  $\text{CO}_2$ ; Supplementary Table 1). Uncertainties, indicated by the widths of the data bands, are defined in the respective studies (see Methods for extra references). The two dashed lines indicate the preindustrial (280 p.p.m.) and current (390 p.p.m. in 2010) concentrations of atmospheric  $\text{CO}_2$ .

further comparison between the mean and ‘interglacial’ trends supports our main conclusions (Methods and Supplementary Figs 1–3).

### Warm pool temperatures and $\text{CO}_2$

The first notable aspect of the Pliocene–Pleistocene climate evolution is the stability of warmest SSTs in the tropical warm pool over the past 5 Myr. In all three tropical oceans, these temperatures stayed fairly constant and the long-term average remained  $\sim 29^\circ\text{C}$  (Figs 2a and 3a), similar to the present. This is especially surprising given temperature changes elsewhere. It has been suggested that Mg/Ca-based measurements, providing many warm pool records, should be corrected for secular changes in seawater composition<sup>21</sup>. However, general agreement between Mg/Ca and alkenone data in locations where both proxies are available (ref. 12 and Methods) indicates that this correction should be small, comparable to calibration errors<sup>12</sup>.

Although a number of thermostat mechanisms capable of maintaining the stability of warm pool SST in high- $\text{CO}_2$  climates have been proposed<sup>22</sup>, we find that they may not be necessary to explain the Pliocene climate. A pervasive increase in atmospheric  $\text{CO}_2$  is the driver for present anthropogenic climate change, and is expected to cause even larger changes in the future<sup>16</sup>. Similarly,  $\text{CO}_2$  remains a suspected cause for early Pliocene warmth, and several biogeochemical methods have been devised to estimate its concentration (Fig. 1c). Despite large uncertainties, together these proxy data suggest that Pliocene  $\text{CO}_2$  concentrations were only 50–100 p.p.m. higher than preindustrial values ( $\sim 280$  p.p.m.). Climate models produce a  $\sim 1^\circ\text{C}$  temperature rise in the warm pool when the  $\text{CO}_2$  concentration increases by 100 p.p.m.<sup>16</sup>, and a change by  $\sim 2^\circ\text{C}$  in doubling-of- $\text{CO}_2$  experiments<sup>23</sup>. Thus, within data uncertainty (Methods), the

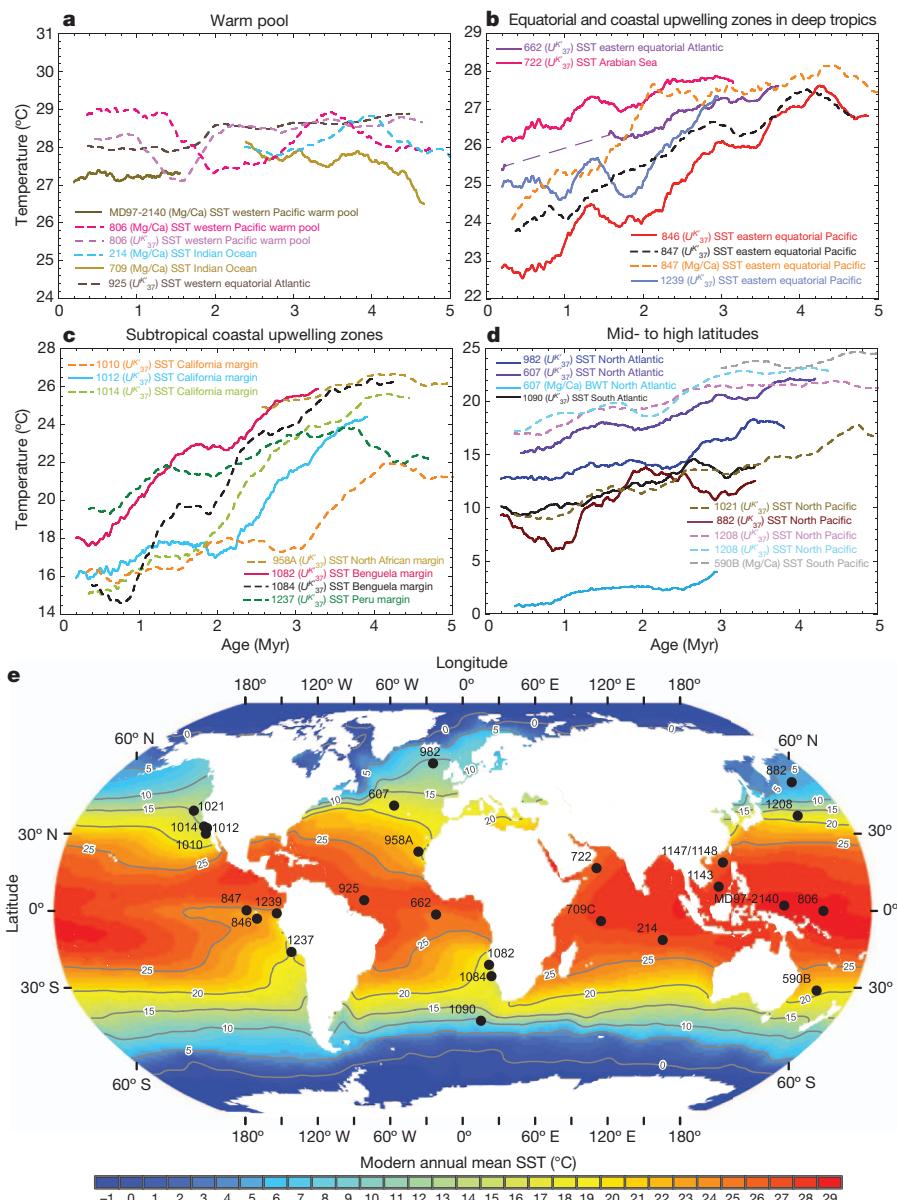
stability of the warm pool temperatures is not inconsistent with the relatively small  $\text{CO}_2$  change.

### Increasing meridional temperature gradients

High-latitude warmth and a reduced Equator-to-pole temperature gradient are other dominant features of early Pliocene climate<sup>2,24</sup>. Marine data, available between latitudes  $43^\circ\text{S}$  and  $58^\circ\text{N}$  (Fig. 2e), suggest subsequent ocean cooling of  $4\text{--}8^\circ\text{C}$  in the mid and high latitudes of the Atlantic and Pacific (Figs 2d and 3d). The temperature of Atlantic deep waters (ODP site 607; 3,400-m depth) basically follows SST evolution in the North Atlantic, implying a similar cooling over the ocean convection regions at higher latitudes ( $\sim 70^\circ\text{N}$ ). In terrestrial regions of the Arctic, temperature fell by nearly  $19^\circ\text{C}$  (ref. 19).

The meridional SST gradient as measured from the Equator to the subtropics was also significantly smaller in the early Pliocene than at present<sup>10</sup> (Fig. 4a–c), and the meridional temperature distribution within the Tropics was more uniform<sup>10</sup>. Consequently, despite little difference in the warmest SSTs (Figs 2a and 3a), the meridional extent of the warm pool was much broader in the early Pliocene. The subsequent cooling led to a gradual contraction of the warm pool towards the Equator, as evidenced by the increase in meridional temperature gradient. This contraction is also apparent from temperature records at the edge of the warm pool in the South China Sea<sup>25,26</sup> (Supplementary Fig. 9). Whereas data from inside the warm pool show fairly constant temperatures, these particular records show a clear cooling trend.

The reduced meridional SST gradient seems to be important for understanding the high global mean temperature and other characteristics of the early Pliocene<sup>8,10</sup>. For instance, from the atmospheric perspective, as the temperature contrast from the Equator to the subtropics increased, the



**Figure 2 | Temperature evolution over the past 5 Myr in different regions of the ocean.** **a–d**, Trends computed as running means of the original proxy temperature data. Solid lines are 400-kyr running means for records that resolve Milankovitch (orbital) timescales, whereas dashed lines are 600-kyr running means of time series with a coarser resolution. Before computing the running means, all time series were interpolated to a constant spacing of 2 kyr. **e**, Map of the sites (black dots) for which the data were generated, superimposed on a map of modern annual mean SST (contours and colours). Numbers indicate particular sites from the Deep Sea Drilling Project (DSDP), the Ocean Drilling Program (ODP), the Integrated Ocean Drilling Program (IODP), and the International Marine Past Global Study (IMAGES). Note that ODP site 607 has records for both surface (SST) and bottom (BWT) water temperature. The original data along with the relevant references are provided in Supplementary Information.

surface high-pressure zones in low latitudes (subtropical highs) strengthened, whereas the Hadley cells (atmospheric meridional circulation) intensified and contracted slightly towards the Equator. The strengthening of atmospheric circulation led to a stronger subsidence, resulting in aridification of parts of Africa, Australia and North America<sup>8,27</sup>.

From the oceanic perspective, as the meridional SST gradient increased so did upper-ocean stratification in the tropics, because oceanic vertical thermal structure is directly related to the meridional distribution of surface temperature<sup>28,29</sup>. This increase in stratification should have contributed to the basin-wide shoaling of the tropical thermocline<sup>9,30,31</sup>, which has been under way since 5 Myr ago<sup>11,13,30,31</sup>.

### Strengthening of cold upwelling

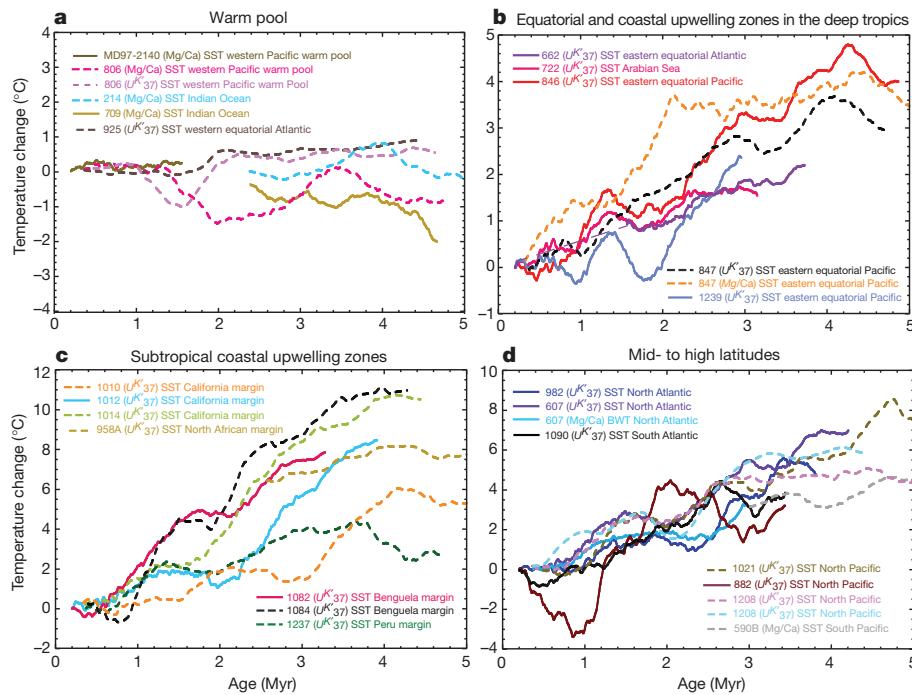
As the meridional SST gradient increased and the tropical thermocline shoaled, cold waters became present at low latitudes, with dramatic consequences for the tropics. Initially the cold waters appeared in the coastal subtropical upwelling sites in the Atlantic and Pacific oceans, in both hemispheres (Fig. 2c and 3c). The strengthening of anticyclonic winds within the subtropical highs may have contributed to the stronger upwelling<sup>8,10</sup>.

The shoaling of the thermocline culminated in the appearance of colder water along the Equator (Figs 2b and 3b) and the formation of

a salient feature of the present-day SST pattern, namely the equatorial ‘cold tongues’ on the eastern sides of the Pacific and Atlantic basins. The corresponding development of the zonal temperature gradients along the Equator (Fig. 4d–f) and the ensuing Bjerknes feedback (a stronger SST gradient leads to stronger winds and, thence, a stronger gradient) led to the intensification of the zonal atmospheric circulation—the Walker cell<sup>3,10</sup>.

In the present climate, an intermittent weakening of both the winds and the east–west equatorial temperature gradient occurs during El Niño, but the mean climate is still characterized by pronounced cold tongues in the Pacific and the Atlantic. These tongues are maintained by a persistent supply of colder water from the extratropics by shallow overturning cells<sup>32</sup>. The cold tongues and subtropical coastal upwelling zones are the main low-latitude regions where the ocean gains heat to balance that lost in high latitudes<sup>27,28</sup>. Apparently, during the early Pliocene the cold water was mainly absent in the eastern equatorial Pacific, and the zonal SST gradient along the Equator was small (Fig. 4d–f). This state is sometimes referred to as ‘permanent El Niño-like’<sup>11</sup> (although this implies nothing about interannual variability). A similar situation occurred in the Atlantic (Fig. 2a, b).

The zonal SST gradient in the two oceans reached its modern values between 2 and 1 Myr ago (Fig. 4). This was a gradual process rather than



**Figure 3 | Temperature changes in different regions of the ocean.** Relative changes in temperature with respect to the beginning of each trend line. Solid and dashed lines are as in Fig. 2. Note the stability of warm pool temperatures (a) with respect to SST changes in other regions of the ocean (b–d). The strongest temperature change is seen in some of the coastal upwelling regions. For changes computed only for interglacial temperatures, see Supplementary Fig. 2.

a swift transition. The magnitude of cooling of the eastern equatorial Pacific ( $4\text{--}5^{\circ}\text{C}$  over 4 Myr) appears to be just slightly lower than that in mid latitudes (Fig. 3b, d).

A zonal SST gradient also developed in the Indian Ocean, where seasonal coastal upwelling became prominent in the Arabian Sea (Figs 2b and 3b). The cooling in this region since the early Pliocene may have contributed to a gradual strengthening of the Asian monsoon through coupling between the winds and SSTs (stronger winds lead to a greater SST contrast and, in turn, yet stronger winds). In an atmospheric general circulation model (GCM), transitioning from a uniformly warm Indian Ocean to a cooler Arabian Sea leads to the strengthening of the monsoon and the aridification of East Africa<sup>8</sup>.

### Mechanisms to explain Pliocene climate

Simulations with atmospheric GCMs forced with different Pliocene SST reconstructions suggest that the global mean temperature was  $2\text{--}3^{\circ}\text{C}$  higher<sup>33</sup> in the mid Pliocene ( $\sim 3$  Myr ago) and roughly  $4^{\circ}\text{C}$  higher<sup>8</sup> in the early Pliocene ( $\sim 4$  Myr ago) than the preindustrial temperature. Many studies have focused on the related problem of how the Northern Hemisphere stayed almost free of continental ice before the onset of the ice ages<sup>25,34,35</sup>. Several hypotheses were proposed, ranging from high CO<sub>2</sub> to enhanced ocean poleward heat transport<sup>2,24</sup> caused by changes in ocean gateways<sup>5,35</sup> and to changes in atmospheric processes and feedbacks.

However, the Pliocene problem<sup>9</sup> goes much farther than the issues of mean global temperatures or continental ice sheets. Accordingly, here we conduct experiments with a state-of-the-art coupled model (Methods) to explore whether these prevailing hypotheses are consistent with the three observational constraints on changes in tropical climate between the present and the early Pliocene, that is, the reduction of meridional temperature gradient by  $4\text{--}6^{\circ}\text{C}$  between the Equator and, respectively,  $30^{\circ}\text{N}$  and  $30^{\circ}\text{S}$ ; a similar  $4\text{--}6^{\circ}\text{C}$  reduction of zonal SST gradient in the equatorial Pacific; and little SST increase ( $0\text{--}1^{\circ}\text{C}$ ) in the warm pool.

### Higher atmospheric CO<sub>2</sub>

In the first experiment, we increase CO<sub>2</sub> by 100 p.p.m. starting from the preindustrial climate. As expected, the climate response to this increase is amplified at high latitudes, leading to a reduction of the Equator-to-pole temperature gradient. Nevertheless, this reduction

between the Equator and mid latitudes reaches only  $0.5^{\circ}\text{C}$  (Fig. 5a). The eastern equatorial Pacific warms by the same amount, but so does the warm pool. As a result, the zonal gradients hardly change, whereas the tropical thermocline deepens very slightly (Fig. 5f). Typically, in other models<sup>36</sup>, even for CO<sub>2</sub> doubling, changes in this gradient do not exceed  $\pm 0.5^{\circ}\text{C}$ , suggesting that CO<sub>2</sub> alone cannot explain the early Pliocene SST patterns.

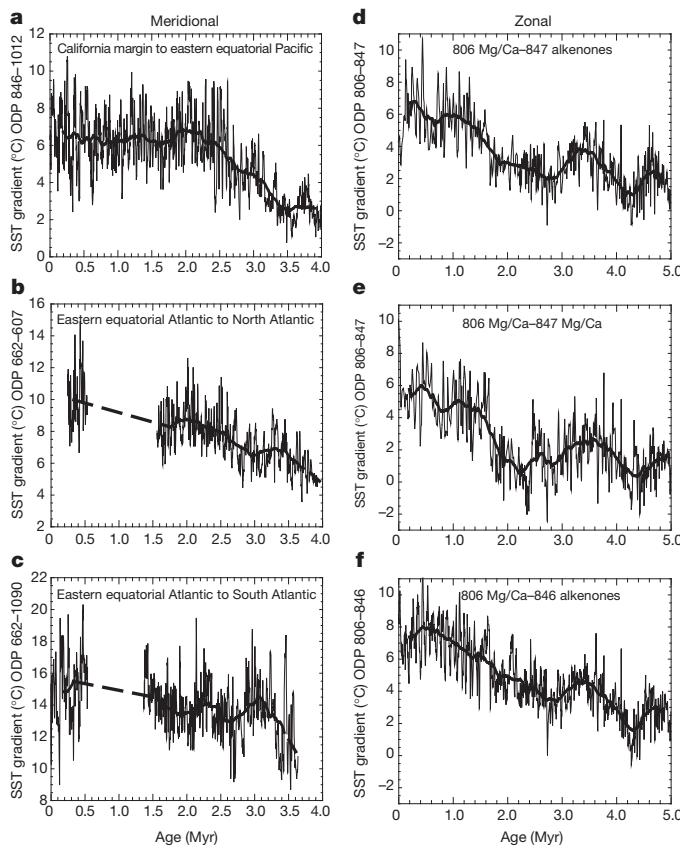
A comprehensive simulation<sup>37</sup> of the mid Pliocene forced by CO<sub>2</sub> concentrations of 400 p.p.m. and PRISM3 land surface boundary conditions shows an increase in global mean temperature of around  $3^{\circ}\text{C}$ , of which half comes from the CO<sub>2</sub> increase and the rest from lower orography and reduced land albedo. The model simulates a warming by  $3\text{--}4^{\circ}\text{C}$  in some regions of the high-latitude Atlantic, partly attributed to the effect of a smaller Greenland ice sheet<sup>15,33,34</sup>; however, SST gradients at low latitudes change little while temperature in the warm pool rises by  $2\text{--}3^{\circ}\text{C}$ . This confirms that by itself the relatively minor increase in CO<sub>2</sub> (50–100 p.p.m.) could not have induced the observed structural climate change, even with high Earth-system climate sensitivity<sup>37,38</sup>.

### Ocean gateways

Enhanced ocean poleward heat transport during the Pliocene has been proposed as another factor that contributes to keeping high latitudes warm, especially in the northern Atlantic<sup>2,24</sup>. However, climate models calculate changes in ocean heat transport interactively and so require particular forcing mechanisms, such as an open Central American seaway<sup>31,39</sup> (CAS).

Originally, the CAS closure was discussed in the context of the onset of glaciation around 2.7 Myr ago<sup>39</sup>, but the closure is now estimated to have occurred between 4.7 and 4.2 Myr ago, as inferred from the divergence of planktonic δ<sup>18</sup>O between the Caribbean Sea and the Pacific<sup>5,31,40</sup>. Recently, it has been suggested that the seaway closure, and the resulting intensification of deep-ocean circulation due to increased salinity in the Atlantic, could have led to a shoaling of the thermocline in the Pacific, facilitating the subsequent cooling<sup>31</sup>.

Opening the CAS in our model to a depth of 150 m (ref. 5) causes a reduction in the strength of the Atlantic meridional overturning circulation, a cooling of the Northern Hemisphere and a warming of the Southern Hemisphere (Fig. 5c, h and Supplementary Fig. 11). However, impacts on the equatorial SSTs are minor. The equatorial



**Figure 4 | Evolution of SST gradients since 4–5 Myr ago.** **a–c**, Meridional SST gradients (in the Pacific and Atlantic) from alkenone SST proxies. **d–f**, Zonal SST gradient along the Equator in the Pacific from different ODP sites and temperature proxies. Despite inherent uncertainties, the different data support the notion of the reduced zonal and meridional SST gradients in the early Pliocene.

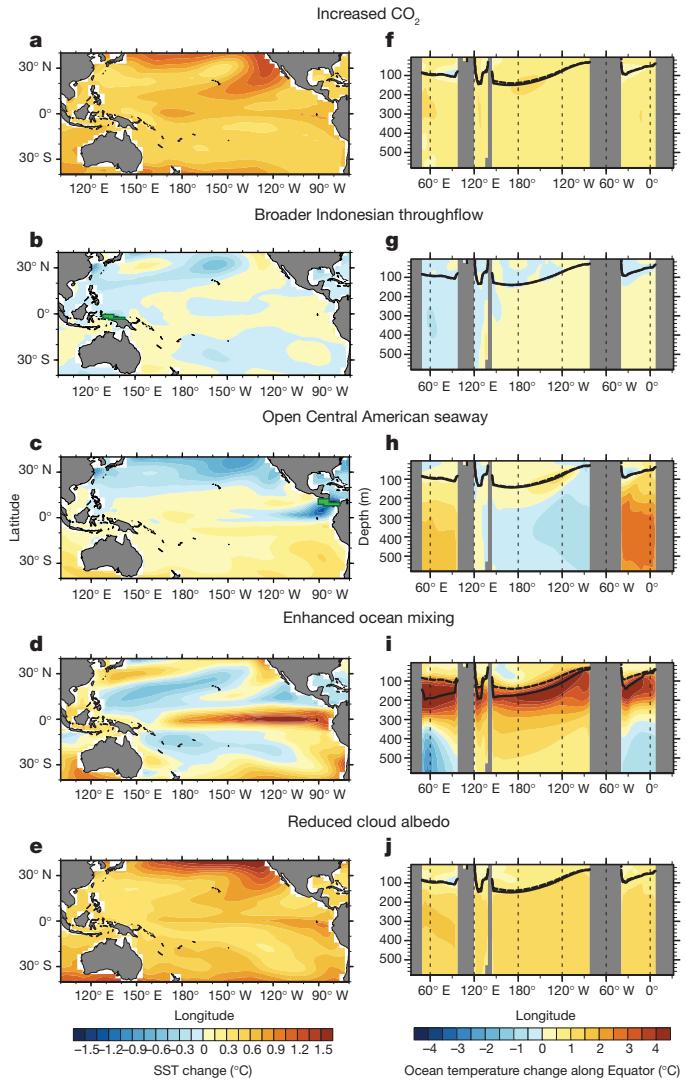
thermocline deepens a little in the central Pacific but shoals in the very east. Opening the CAS to the largest possible depth (1,100 m) barely changes the zonal SST gradient along the Equator, but amplifies the interhemispheric seesaw pattern with a strong cooling in the northern high latitudes (Supplementary Fig. 11), contradicting the observations. Similar behaviour is seen in other models (ref. 41; Supplementary Fig. 12).

Another tectonic factor potentially important for Pliocene climate is the northward drift of New Guinea, which may have altered the source of waters entering the Indonesian throughflow and feeding the Equatorial undercurrent and upwelling<sup>42</sup>. However, in agreement with previous modelling studies<sup>43</sup>, our results show impacts barely above statistical significance (Fig. 5b, g) and do not explain the Pliocene climate state.

#### Effects of ocean mixing and cloud albedo

Recently, dynamical feedbacks not included or possibly underestimated in climate models have also been suggested to explain early Pliocene warmth<sup>44,45</sup>. In particular, it has been argued that tropical cyclones during that time were more widespread, generating strong ocean vertical mixing in the subtropical bands. Such enhanced mixing would reduce oceanic heat transport from the equatorial region to the subtropics, deepening the tropical thermocline and warming the cold tongue. Presumably, the contraction of the warm pool and associated strengthening of vertical wind shear decreased tropical cyclone activity, contributing to the cooling trends<sup>45</sup>.

Our experiments show that a plausible increase in the subtropical mixing (Supplementary Fig. 13b) causes a reduction of the zonal SST



**Figure 5 | Testing mechanisms of Pliocene warmth and reduced temperature gradients.** Anomalies in SST (a–e) and subsurface temperature (f–j) between the simulated early Pliocene and the preindustrial control. All simulations use a coupled climate model (the NCAR Community Earth System Model) and are identical except for a particular model perturbation: increasing  $\text{CO}_2$  concentrations by 100 p.p.m. from the preindustrial level (a, f); shifting islands of Indonesia southward (b, g); opening the CAS to a depth of 150 m (c, h); increasing ocean mixing in the subtropical bands (d, i); and reducing extratropical cloud albedo (e, j). Landmasses blocked out in green were removed in the respective experiments. Thick dashed and solid lines in the right panels show the thermocline depth in the Control and perturbation experiments, respectively.

gradient in the Pacific by  $\sim 2^\circ\text{C}$ , which is promising (Fig. 5d, i). This is the only experiment that causes a broad deepening of the thermocline along the Equator (Fig. 5i). However, the impacts also include an increase by almost  $1^\circ\text{C}$  in the meridional SST gradient (associated with ocean cooling in hurricane wakes), rather than the required reduction.

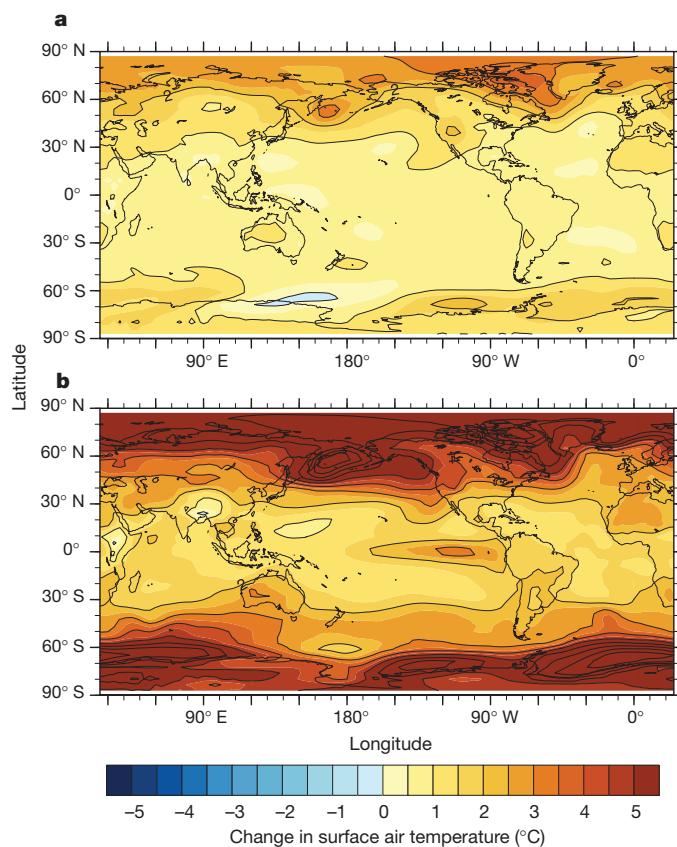
It has also been suggested that extratropical low stratus clouds were less reflective in the Pliocene<sup>44</sup>, possibly as a result of a different aerosol composition or a stronger cloud response to surface warming than coupled models typically predict. Our sensitivity experiment with a moderately reduced low-cloud albedo (Supplementary Fig. 13a) shows a reduction by 1–2  $^\circ\text{C}$  in the meridional SST gradient, a weak reduction in the zonal gradient and a modest deepening of the tropical thermocline (Fig. 5e, j). However, somewhat like the  $\text{CO}_2$  experiment, this simulation increased warm pool SSTs while producing little change in the thermocline depth.

## Quantifying structural climate change

We can assess how well these experiments capture a key characteristic of the Pliocene climate, that is, a more uniform SST field at low latitudes. A quantitative measure that reflects the structural aspects of climate change is the change in the spatial variance of time-averaged temperatures evaluated for particular regions of the ocean (Supplementary Table 4). Weakened zonal and meridional temperature gradients imply less spatial diversity in temperature, and so the early Pliocene should have a smaller variance at low latitudes as compared with the present. None of the aforementioned simulations reduces the tropical variance computed within the band between 30° N and 30° S by more than 5%, whereas some actually increase it. When computed for the equatorial band (10° N to 10° S), the variance is 40% lower in the ocean-mixing experiment but less than 15% lower in the others.

## Discussion and outlook

Palaeorecords allow us to identify critical features of the early Pliocene climate: it was a warm climate characterized by a minimal increase in warm pool SST but substantially weaker meridional and zonal SST gradients, and, hence, weaker atmospheric circulation<sup>3,8,10,45</sup>, and a deeper tropical thermocline. These differences between the early Pliocene and modern climates amount to structural climate change, regardless of whether we consider trends in mean temperatures or interglacial temperatures (the latter are less affected by continental ice sheets than are the former).



**Figure 6 | Global surface air temperature anomalies in two experiments.** **a**, CO<sub>2</sub> increase; **b**, Combined. The Combined experiment includes three different forcings: a 100-p.p.m. CO<sub>2</sub> increase, enhanced ocean mixing in the subtropical bands and reduced albedo of low clouds in the extratropics. Note the warming of the eastern equatorial Pacific, a reduction of the temperature gradient between the Equator and the subtropics (also see Supplementary Fig. 11), and a substantial warming of high latitudes.

Overall, we are just beginning to understand the driving forces behind climate evolution since the early Pliocene. The climatic changes occurred in concert with a relatively small reduction—less than 50–100 p.p.m.—in atmospheric CO<sub>2</sub> (Fig. 1c). Either this reduction pushed the system over a threshold, leading to the structural climate change, or the climate system was cooled by other mechanisms and CO<sub>2</sub> provided a positive feedback for the cooling.

None of the currently proposed mechanisms can reproduce all of the key Pliocene climate features by itself. Some mechanisms address one aspect of the structural change, but often at the detriment of the other features; a major problem is to simulate reduced temperature gradients without raising warm pool SSTs. Increases in greenhouse gases alone are not sufficient, whereas shifts in oceanic gateways have much smaller effects than sometimes suggested. The most promising dynamical mechanisms involve lower extratropical cloud albedo or enhanced ocean mixing. For these mechanisms to explain the stark differences between the Pliocene and today's climates, the corresponding model has to be modified.

Imposing the two dynamical mechanisms in a climate model while increasing CO<sub>2</sub> (Fig. 6b) produces a warming by ~3 °C in the eastern equatorial Pacific and a reduction by 2–4 °C in the meridional temperature gradient, leading to an increase of only 1 °C in the temperature of the warm pool. The spatial variance of SSTs decreases by 60% in the equatorial region (Supplementary Table 4), in stark contrast with the CO<sub>2</sub>-only experiment (Fig. 6a). Thus, a combination of several factors may be necessary to explain Pliocene climate.

An important question is to what degree the evidence of climate evolution over the last 5 Myr shapes our assessment of future change. From the observations, it is clear that the climate system is capable of remarkable transformations in response even to small changes in external parameters (CO<sub>2</sub> concentrations, continental boundaries and orography are generally similar for the early Pliocene and present). Therefore, explaining the discrepancy between model simulations and the early Pliocene temperature patterns is essential for building confidence in our climate projections. Finally, it might be necessary to contemplate, in addition to global mean temperature increase, the possibility of structural climate change (relaxation of prevailing temperature gradients at low latitudes).

## METHODS SUMMARY

All the geochemical records used to compute the trends presented in this article have been previously published. The ocean temperatures come from two geochemical proxies (alkenone unsaturation index ( $U^{37}_K$ ) and Mg/Ca ratios of planktonic foraminiferal shells) that are derived from material preserved in deep-sea sediment (Methods). These proxies are recorded by microorganisms living in the surface mixed layer of the ocean (except one record that comes from benthic foraminifers), where their chemical composition reflects the conditions of the surrounding water. The leading-order trends are computed as 400-kyr or 600-kyr running means for each temperature record, which filters out glacial-interglacial variability. The original data are included in Supplementary Information. The CO<sub>2</sub> estimates were calculated using a variety of techniques (Supplementary Table 1) and are presented as originally published except the record based on  $\delta^{13}\text{C}$  (ref. 46), which was adjusted to subtract an offset (90 p.p.m.) between the original data and the Antarctic ice-core CO<sub>2</sub> records.

The climate model used is the NCAR Community Earth System Model<sup>47</sup>. This incorporates an atmospheric model and a land surface model (CAM4 and CLM, respectively) with a spectral truncation of T31 (a horizontal grid of 3.75° by 3.75°) coupled to ocean and sea-ice components (POP2 and CICE) with a nominal 3° resolution that increases to 1° near the Equator. Each numerical experiment is designed to examine the effects of a single type of perturbation with respect to a preindustrial baseline simulation (a change in a particular ocean gateway or a change in cloud properties, for example). In the final experiment, titled 'Combined', we imposed three different perturbations together (enhanced ocean mixing in the subtropical bands, reduced albedo of low extratropical clouds and a 100-p.p.m. CO<sub>2</sub> increase). For each experiment, the simulation time was 500 yr, with the final 25 yr being used for analysis. Further details are provided in Methods, including a discussion of mechanisms related to ocean salinity<sup>48</sup> and atmospheric superrotation<sup>49</sup>.

**Full Methods** and any associated references are available in the online version of the paper.

Received 16 October 2012; accepted 11 February 2013.

1. Shackleton, N. J. et al. Oxygen isotope calibration of the onset of ice-raftering and history of glaciation in the North Atlantic region. *Nature* **307**, 620–623 (1984).
2. Crowley, T. Pliocene climates: the nature of the problem. *Mar. Micropaleontol.* **27**, 3–12 (1996).
3. Ravelo, A. C., Andreassen, D. H., Lyle, M., Lyle, A. O. & Wara, M. W. Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature* **429**, 263–267 (2004).
4. Jansen, E., Fronval, T., Rack, F. & Channell, J. E. T. Pliocene-Pleistocene ice rafting history and cyclicity in the Nordic seas during the last 3.5 Myr. *Paleoceanography* **15**, 709–721 (2000).
5. Haug, G., Tiedemann, R., Zahn, R. & Ravelo, A. Role of Panama uplift on oceanic freshwater balance. *Geology* **29**, 207–210 (2001).
6. Ruddiman, W. Orbital insolation, ice volume, and greenhouse gases. *Quat. Sci. Rev.* **22**, 1597–1629 (2003).
7. Crowley, T. & Hyde, W. Transient nature of late Pleistocene climate variability. *Nature* **456**, 226–230 (2008).
8. Brierley, C. & Fedorov, A. Relative importance of meridional and zonal sea surface temperature gradients for the onset of the ice ages and Pliocene-Pleistocene climate evolution. *Paleoceanography* **25**, PA2214 (2010).
9. Fedorov, A. et al. The Pliocene paradox (mechanisms for a permanent El Niño). *Science* **312**, 1485–1489 (2006).
10. Brierley, C. M. et al. Greatly expanded tropical warm pool and weakened Hadley circulation in the early Pliocene. *Science* **323**, 1714–1718 (2009).
11. Wara, M. W., Ravelo, A. C. & Delaney, M. L. Permanent El Niño-like conditions during the Pliocene warm period. *Science* **309**, 758–761 (2005).
12. Dekens, P., Ravelo, A., McCarthy, M. & Edwards, C. A 5 million year comparison of Mg/Ca and alkenone paleothermometers. *Geochem. Geophys. Geosyst.* **9**, Q10001 (2008).
13. Steph, S., Tiedemann, R., Groeneveld, J., Sturm, A. & Nurnberg, D. Pliocene changes in tropical east Pacific upper ocean stratification: Response to tropical gateways? *Proc. Ocean Drill. Program Sci. Results* **202**, 1–51 (2006).
14. Lawrence, K., Herbert, T., Brown, C., Raymo, M. & Haywood, A. High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanography* **24**, PA2218 (2009).
15. Sosdian, S. & Rosenthal, Y. Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* **325**, 306–310 (2009).
16. Solomon, S. et al. *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
17. Dowsett, H. J. et al. *Middle Pliocene Paleoenvironmental Reconstruction: PRISM2*. Open File Report 99–535 (US Geol. Surv., 1999).
18. Dowsett, H. J. et al. The PRISM3D paleoenvironmental reconstruction. *Stratigraphy* **7**, 123–139 (2010).
19. Ballantyne, A. et al. Significantly warmer arctic surface temperatures during the Pliocene indicated by multiple independent proxies. *Geology* **38**, 603–606 (2010).
20. Marlow, J., Lange, C., Wefer, G. & Rosell-Mele, A. Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science* **290**, 2288–2291 (2000).
21. Medina-Elizalde, M., Lea, D. W. & Fantle, M. S. Implications of seawater Mg/Ca variability for Plio-Pleistocene tropical climate reconstruction. *Earth Planet. Sci. Lett.* **269**, 585–595 (2008).
22. Williams, I., Pierrehumbert, R. & Huber, M. Global warming, convective threshold and false thermostats. *Geophys. Res. Lett.* **36**, L21805 (2009).
23. DiNezio, P. N. et al. Climate response of the equatorial pacific to global warming. *J. Clim.* **22**, 4873–4892 (2009).
24. Chandler, M., Rind, D. & Thompson, R. Joint investigations of the middle Pliocene climate II: GISS GCM Northern Hemisphere results. *Global Planet. Change* **9**, 197–219 (1994).
25. Jia, G., Chen, F. & Peng, P. Sea surface temperature differences between the western equatorial Pacific and northern South China Sea since the Pliocene and their paleoclimatic implications. *Geophys. Res. Lett.* **35**, L18609 (2008).
26. Li, L. et al. A 4-Ma record of thermal evolution in the tropical western Pacific and its implications on climate change. *Earth Planet. Sci. Lett.* **309**, 10–20 (2011).
27. deMenocal, P. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth Planet. Sci. Lett.* **220**, 3–24 (2004).
28. Boccaletti, G., Pacanowski, R., Philander, S. & Fedorov, A. The thermal structure of the upper ocean. *J. Phys. Oceanogr.* **34**, 888–902 (2004).
29. Fedorov, A. V., Pacanowski, R. C., Philander, S. G. & Boccaletti, G. The effect of salinity on the wind-driven circulation and the thermal structure of the upper ocean. *J. Phys. Oceanogr.* **34**, 1949–1966 (2004).
30. Ford, H. L., Ravelo, A. C. & Hovan, S. A deep Eastern Equatorial Pacific thermocline during the early Pliocene warm period. *Earth Planet. Sci. Lett.* **355–356**, 152–161 (2012).
31. Steph, S. et al. Early Pliocene increase in thermohaline overturning: a precondition for the development of the modern equatorial pacific cold tongue. *Paleoceanography* **25**, PA2202 (2010).
32. McCreary, J. & Lu, P. Interaction between the subtropical and equatorial ocean circulations: the subtropical cell. *J. Phys. Oceanogr.* **24**, 466–497 (1994).
33. Haywood, A. M. et al. Comparison of mid-Pliocene climate predictions produced by the HADAM3 and GCMAM3 general circulation models. *Global Planet. Change* **66**, 208–224 (2009).
34. Lunt, D., Foster, G., Haywood, A. & Stone, E. Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO<sub>2</sub> levels. *Nature* **454**, 1102–1105 (2008).
35. Maier-Reimer, E., Mikolajewicz, U. & Crowley, T. Ocean general circulation model sensitivity experiment with an open Central American isthmus. *Paleoceanography* **5**, 349–366 (1990).
36. Collins, M. El Niño or La Niña-like climate change? *Clim. Dyn.* **24**, 89–104 (2005).
37. Lunt, D. et al. Earth system sensitivity inferred from Pliocene modelling and data. *Nature Geosci.* **3**, 60–64 (2010).
38. Pagan, M., Liu, Z., LaRiviere, J. & Ravelo, A. High earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nature Geosci.* **3**, 27–30 (2010).
39. Haug, G. & Tiedemann, R. Effect of the formation of the Isthmus of Panama on Atlantic ocean thermohaline circulation. *Nature* **393**, 673–676 (1998).
40. Molnar, P. Closing of the Central American seaway and the ice age: a critical review. *Paleoceanography* **23**, PA2201 (2008).
41. Zhang, X. et al. Changes in equatorial Pacific thermocline depth in response to Panamanian seaway closure: insights from a multi-model study. *Earth Planet. Sci. Lett.* **317–318**, 76–84 (2012).
42. Cane, M. A. & Molnar, P. Closing of the Indonesian seaway as a precursor to east African aridification around 3–4 million years ago. *Nature* **411**, 157–162 (2001).
43. Jochum, M., Fox-Kemper, B., Molnar, P. & Shields, C. Differences in the Indonesian Seaway in a coupled climate model and their relevance to Pliocene climate and El Niño. *Paleoceanography* **24**, PA1212 (2009).
44. Barreiro, M. & Philander, S. G. Response of the tropical Pacific to changes in extratropical clouds. *Clim. Dyn.* **31**, 713–729 (2008).
45. Fedorov, A., Brierley, C. & Emanuel, K. Tropical cyclones and permanent El Niño in the early Pliocene epoch. *Nature* **463**, 1066–1070 (2010).
46. Raymo, M., Grant, B., Horowitz, M. & Rau, G. Mid-Pliocene warmth: stronger greenhouse and stronger conveyor. *Mar. Micropaleontol.* **27**, 313–326 (1996).
47. Gent, P. R. et al. The Community Climate System Model version 4. *J. Clim.* **24**, 4973–4991 (2011).
48. Fedorov, A., Barreiro, M., Boccaletti, G., Pacanowski, R. & Philander, S. The freshening of surface waters in high latitudes: effects on the thermohaline and wind-driven circulations. *J. Phys. Oceanogr.* **37**, 896–907 (2007).
49. Tziperman, E. & Farrell, B. Pliocene equatorial temperature: lessons from atmospheric superrotation. *Paleoceanography* **24**, PA1101 (2009).
50. Lisiecki, L. E. & Raymo, M. E. A. Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, 1–17 (2005).
51. Haug, G. et al. North Pacific seasonality and the glaciation of North America 2.7 million years ago. *Nature* **433**, 821–825 (2005).
52. Shipboard. Scientific Party in *Proc. ODP Init. Reports*. Vol. 162 (eds Jansen, E. et al.) 91–138 (Texas A & M Univ., 1996).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Financial support was provided by the grants to A.V.F. and K.T.L. from the US National Science Foundation (OCE-0901921 and OCE-0623310), the US Department of Energy Office of Science (DE-SC0007037) and the David and Lucile Packard Foundation. Most of the data compiled in this study comes from samples collected by the Deep Sea Drilling Project, the Ocean Drilling Program, and the Integrated Ocean Drilling Program, supported by the National Science Foundation. This work was supported in part by the Yale University Faculty of Arts and Sciences High Performance Computing facility. A.V.F. thanks G. Philander, P. deMenocal, T. Herbert, Y. Rosenthal, A. Haywood, D. Lunt, M. Barreiro, E. Tziperman, D. Battisti, M. Cane and C. Wunsch for discussions of this topic. We are grateful to a great number of scientists who provided the data used in this study.

**Author Contributions** A.V.F., K.T.L. and C.M.B. coordinated the manuscript and contributed equally to writing the manuscript and to the ideas it contains; K.T.L. compiled the SST data with Z.L. (alkenones) and P.S.D. (Mg/Ca). C.M.B. performed the model simulations with A.V.F.; A.V.F. compiled the CO<sub>2</sub> data. A.C.R. helped to interpret the data. All authors were involved in the writing at different stages of the work on this manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.V.F. (alexey.fedorov@yale.edu).

## METHODS

**SST reconstructions.** Two geochemical proxies, alkenone unsaturation index ( $U^{K_{37}}$ ) and Mg/Ca ratios of planktonic foraminiferal shells, derived from material preserved in deep-sea sediment, are now widely used to reconstruct ocean temperatures. Alkenones are long-chain ethyl and methyl ketones that vary in the number of double bonds they have and are produced by coccolithophorids living in ocean surface waters. The degree of unsaturation in  $C_{37}$  alkenones reflects the temperature of the water in which these organisms lived at the time of production. Owing to their large size relative to other organic compounds, these molecules are resistant to degradation and are therefore well preserved in the sediment record. Several global calibration studies indicate a robust linear calibration between  $U^{K_{37}}$  and ocean surface temperature, confirming the global applicability of a single calibration equation (see ref. 53 for a review). All  $U^{K_{37}}$  values reported here are use the calibration from ref. 54 except for ODP site 882, which, owing to a significant abundance of  $C_{37:4}$ , uses the  $U^{K_{37}}$  index as initially reported<sup>55</sup>. Because the dominant producers of  $U^{K_{37}}$  in the modern ocean have been present only in the past 200 kyr, there was initially concern about the applicability of this proxy on longer timescales. However, numerous studies have shown that changes in  $U^{K_{37}}$  are not coincident with shifts in haptophyte algae species<sup>56,57</sup>, indicating that producers of alkenones consistently record surface water conditions.

Planktonic foraminifera secrete their  $\text{CaCO}_3$  shell in the surface mixed layer, where warmer temperatures cause an exponential increase in the Mg/Ca ratio because more Mg substitutes for Ca in the shell under warmer conditions. Although shells are commonly well preserved in sediments above the lysocline, partial dissolution of the shells lowers the Mg content, potentially biasing SST estimates. This bias can be avoided by selecting well-preserved shell material or by using calibrations that include a correction for dissolution. The calibration used for any given site must be considered individually, as the preservation of the sample material will determine which calibration is most appropriate. We examined the appropriateness of the calibrations used at each site in the original publications, and assessed the sensitivity of the resultant SST estimates to using or not using a dissolution correction where one had not been applied in the original study. In all cases, the differences in calibrations yielded estimates that differed from each other by less than 1 °C. In most cases, the results differed by less than 0.5 °C, suggesting that SST estimates for the sites examined here are not strongly influenced by dissolution.

Four different calibrations were applied to different records in this study (Supplementary Table 3). The main differences between the calibrations are the presence or absence of a dissolution correction and how that dissolution correction is incorporated into the calibration. It is important to note that the calibrations are quite similar if preservation is good. For example, some calibrations use  $\Delta\text{CO}_3^{2-}$  or core depth to correct for dissolution<sup>58</sup>, but if preservation is excellent, and that correction is removed, the calibration is essentially the same as that of ref. 59, which does not include a dissolution correction. Therefore, calibration choice affects the absolute SST estimates but does not change the long-term trends of the records. Given that changing calibrations from those selected in the original publications yields very small differences in the absolute SST estimates and does not affect the long-term trends, we chose to use the originally applied calibrations at all sites except one. The Mg/Ca data from ODP site 709 overlap with data from ODP site 806, which is not unexpected given that both sites are in the warm pool and at similar water depths. However, the SSTs reported in ref. 60 are ~1.5 °C cooler than those from ODP site 806. This could be because the dissolution correction applied to the data from ODP site 709 is based on core tops from the Caribbean Sea<sup>61</sup>, which may not be appropriate for an Indian Ocean site, such as site 709. We chose to apply the same  $\Delta\text{CO}_3^{2-}$ -corrected calibration to ODP site 709 as was used for ODP site 806 because doing so yields a core-top estimate close to modern SST values in the region for ODP site 806 (ODP site 709 does not have core-top data).

It has been suggested that Mg/Ca temperature measurements, which provide the majority of the warm pool data, should be corrected for secular changes in seawater composition<sup>21</sup>. However, the general agreement between Mg/Ca and alkenone records in locations where both techniques are applicable indicates that this correction should be small, of the order of calibration errors<sup>12</sup>, and it is therefore not applied here.

Finally, we have chosen to omit Mg/Ca records from sites 999 and 1000 located in the Caribbean Sea, because these records do not clearly reflect open-ocean conditions<sup>62</sup>. We have also omitted site 1241 from our analysis, because this site has moved considerably because of tectonic processes (the SST at its present location is ~1.4 °C higher than that at its Pliocene latitude). Additionally, the Mg/Ca calibration used in the original publication<sup>63</sup> is based on a culturing study that excludes a dissolution correction, which we feel should be incorporated at this location.

The trends are computed as 400-kyr or 600-kyr running means for each temperature record, which were resampled to a 2-kyr spacing before computing the means (the original data are included in Supplementary Information). This approach filters out glacial–interglacial variability and highlights first-order trends in these records. Comparative analysis of different resampling resolutions and window sizes demonstrates that the first-order results reported here are insensitive to changes in these parameters.

Computing the trends as we do in the main part of the paper, incorporating both glacial and interglacial intervals of the time series, is conventional in palaeoclimatology. Nevertheless, some of the inferences in this study actually concern differences between the late Quaternary interglacials and Pliocene interglacials. Therefore, to make sure that our conclusions about the structural climate change are robust regardless of how we compute the trends, we have compared the mean trends and those based just on interglacial intervals of the time series.

The procedure for calculating ‘interglacial’ trends includes several steps. First, each time series is divided into separate intervals (bins). Typically, we choose bins of 41-kyr length, which corresponds to the dominant spectral component of the glacial cycles of the past 5 Myr (the one related to variations in Earth’s obliquity). Longer intervals are chosen for coarse time series. For each bin, we find the point with the maximum temperature and then interpolate these temperatures on the original grid of the time series. Finally, to compute the new trends, we apply 400-kyr or 600-kyr running means to the new record. This procedure picks up maximum interglacial temperatures as well as points with relatively warm temperatures but lying outside the actual interglacials (specifically, those belonging to the warm interstadials of the past 1 Myr). We choose to retain such points because many of them are as warm as modern temperatures and also because we prefer to use bins of the same size throughout the entire length of the time series. Very rarely does this procedure pick up cold points erroneously, but the effect of such errors is minimized by computing the running means. Using larger bins, 80 kyr for example, would eliminate most of these points and raise the trend lines by a few tenths of a degree Celsius on average.

As expected, the mean temperatures computed for interglacials (Supplementary Fig. 1) turn out to be ~1 °C warmer than those computed for the full time series, but the relative temperature changes (Supplementary Fig. 2) are remarkably similar between the two methods, with differences typically below 0.5 °C, which confirms the robustness of our main conclusions.

The SST data are grouped into four broad categories: waters of low-latitude warm pools<sup>11,38,60,64,65</sup> (Supplementary Fig. 4); upwelling zones in the deep tropics<sup>11,14,66–70</sup>, including equatorial upwelling sites and the Arabian Sea (Supplementary Fig. 5); subtropical coastal upwelling regions<sup>10,20,67,71–73</sup> (Supplementary Fig. 7); mid and high latitudes<sup>14,15,38,55,72,74–76</sup> (Supplementary Fig. 6). There are also two sites in the South China Sea that currently lie outside the warm pool<sup>25,26,67</sup>, but they seem to have been contained within it earlier (Supplementary Figs 8 and 9). The high productivity at upwelling locations makes them especially suitable for the creation of continuous records over the past 5 Myr, hence their prominence in this compilation. See Supplementary Information for both the full and the temporally averaged data.

**Data uncertainty and error bars.** The two SST proxies we use (Mg/Ca and  $U^{K_{37}}$ ) are widely applied in palaeoceanographic studies, but each carries inherent uncertainties. Although some uncertainties, such as ecological shifts for the  $U^{K_{37}}$  values and changes in seawater chemistry for Mg/Ca values, are difficult to quantify, several studies have now demonstrated that these are probably minor and that both proxies are reliable on Pliocene–Pleistocene timescales<sup>12,53</sup>. Also, although the  $U^{K_{37}}$  index is insensitive to temperature change above ~29 °C, only three of the records included here reach this threshold value, and do so only for short durations. Comparison between these records and other SST proxies from the same sites indicates that temperatures during these times were probably only marginally warmer than the threshold value, which should not affect the conclusions reached in this paper. Because analytical precision is excellent for both types of measurements, the dominant source of error in converting the primary data to SST estimates is in the calibration used. The calibration errors can be propagated to provide an estimate of the errors for each type of data.

Although a detailed error analysis of each of the previously published proxy records included in this study is beyond the scope of this paper, we nonetheless compare four records from the tropical Pacific Ocean (Supplementary Fig. 10). The Mg/Ca and  $U^{K_{37}}$  records at ODP site 847 have previously been compared in detail<sup>13</sup>, and here we compare those records from the eastern equatorial Pacific (EEP) with Mg/Ca and  $U^{K_{37}}$  records from ODP site 806 in the western equatorial Pacific<sup>11,38</sup> (WEP). Because the data were not measured on the same samples within each core, and because of the different time resolutions of the data, we separated the data into 400-kyr bins. This bin size allows for reliable statistics but is not so large that any one bin encompasses a large portion of the long-term trends.

For both the Mg/Ca and  $U^{K'}$ <sub>37</sub>, the analytical error is essentially negligible and the dominant source of error is the calibration used to convert the primary data to a SST estimate. To assess the uncertainty in SST estimates ( $\sigma_{SST}$ ), we propagated the uncertainties in the parameters of the proxy calibrations. For example, for  $U^{K'}$ <sub>37</sub> the linear calibration is

$$SST_{U^{K'}_{37}} = \frac{U^{K'}_{37} - a}{b} - \frac{a}{b}$$

The uncertainties in  $a$  ( $\sigma_a$ ) and  $b$  ( $\sigma_b$ ) are reported in the original calibration, and can be propagated for each data point as follows:

$$\sigma_{SST}^2 = \sigma_a^2 \left( \frac{\partial SST}{\partial a} \right)^2 + \sigma_b^2 \left( \frac{\partial SST}{\partial b} \right)^2 + \sigma_{U^{K'}_{37}}^2 \left( \frac{\partial SST}{\partial U^{K'}_{37}} \right)^2$$

The same process was done for the Mg/Ca calibration (the same Mg/Ca calibration was used in the WEP and the EEP). The average propagated errors for  $U^{K'}_{37}$  are 1.6 °C, for ODP sites 847 and 806, and 1.7 °C and 1.9 °C, for ODP sites 847 and 806, respectively. The error bars in Supplementary Fig. 10 represent the uncertainties in the means calculated for each bin, given the propagated errors described above, calculated as

$$\sigma_{\text{mean SST}} = \sigma_{SST} / \sqrt{N}$$

We note that the errors for  $U^{K'}_{37}$  at ODP site 806 are larger owing to the lower resolution data.

Propagating the error in the calibration demonstrates several important points. First, the Mg/Ca and  $U^{K'}_{37}$  records agree remarkably well in both the WEP and the EEP, in terms of both the long-term trend and the absolute temperature estimates. In the EEP, both records show a cooling by 3–4 °C over the last 4 Myr, and the error bars of the most recent three bins do not overlap with the errors of the oldest three bins, indicating that the EEP was significantly colder in the Pleistocene than during the early Pliocene climatic optimum. The situation is strikingly different in the WEP, where both Mg/Ca and  $U^{K'}_{37}$  indicate that SST in the early Pliocene climatic optimum was not significantly warmer than the Pleistocene. Thus, the errors demonstrate that at both sites the two proxies, which have different sources of potential bias, agree remarkably well (Supplementary Fig. 10). This exercise confirms that SST estimates for the eastern and western tropical Pacific indeed overlap during the early Pliocene, although there is a clear gradient between the two locations in the Pleistocene.

In addition, although there are uncertainties in absolute SST values for the Pliocene, planktonic foraminiferal assemblages and the  $\delta^{18}\text{O}$  of surface and subsurface planktonic foraminifera were also similar in the western and eastern equatorial Pacific (refs 30,78 and references therein), providing strong supporting evidence for the reduced east–west gradients in oceanographic conditions during that time.

Note that the isotopic oxygen ratio used in Fig. 1, where it is expressed in per mille, is defined as  $\delta^{18}\text{O}_{\text{PDB}} (\text{‰}) = [({}^{18}\text{O}/{}^{16}\text{O})_{\text{sample}}/({}^{18}\text{O}/{}^{16}\text{O})_{\text{PDB}} - 1] \times 1,000$  (PDB, PeeDee Belemnite).

**Sensitivity experiments.** All of the pertinent hypotheses, some tested in idealized or more comprehensive models before, have been re-examined within a single modelling framework to allow a better comparison of their relative impacts. This framework is based on the latest version of the NCAR Community Earth System Model<sup>47</sup> (CESM) at its relatively low resolution, which facilitates palaeoclimate simulations<sup>79</sup>. The model incorporates an atmospheric model and a land surface model with a spectral truncation of T31 (CAM4 and CLM) coupled to ocean and sea-ice components with a nominal 3° resolution (POP2 and CICE, respectively). The model has a cold bias that is especially pronounced in the Arctic<sup>79</sup>.

The previous version of the model (CCSM3) has been used extensively for climate simulations that contributed, for example, to the IPCC Fourth Assessment Report<sup>16</sup>; simulations with CESM will contribute to the next report. Further information about the model, including its source code and the boundary conditions are freely accessible via <http://www.cesm.ucar.edu/models/cesm1.0>. We use subversion 1.0.2, because earlier versions of the model were not scientifically validated at the low resolution. All the integrations start from a pre-existing preindustrial simulation<sup>79</sup> (available as b40.037 from <http://www.cesm.ucar.edu/models/cesm1.0>) and have been integrated for a further 500 yr (including the preindustrial control). This is sufficient for the upper ocean to reach equilibrium with the deep ocean experiencing a weak residual drift. All simulations have orography, vegetation type and land-ice cover prescribed at preindustrial conditions and are designed to assess the roles of separate mechanisms rather than to reproduce the early Pliocene climate exactly. Previous work<sup>37</sup> shows that modifying these factors as appropriate for the Pliocene may increase global mean temperature by ~1 °C, but contributes little to the structural climate change

described here. Figures 5 and 6 show the difference in the averaged climate over the final 25 years from this control simulation.

Each numerical experiment examines the effects of a single type of perturbation with respect to the preindustrial baseline simulation (control), except for the final experiment, which combines three different perturbations. Supplementary Table 4 lists all the experiments and the simulated changes in global mean surface air temperature and temperature variances. The other details of the simulations are described below.

The CO<sub>2</sub> experiment involves an instantaneous increase in atmospheric CO<sub>2</sub> concentration from 285 p.p.m. to 385 p.p.m., leading to an increase of 1 °C in global mean air temperature. It is important that the base model has an equilibrium climate sensitivity of 2.9 °C (ref. 80), in the middle of the typical range of model climate sensitivities.

Bathymetric perturbations in the ocean gateway experiments involve the removal of five land points (coloured green in Fig. 5). In the shallow-CAS experiment, the bathymetry has been lowered to 150 m. In the other experiments, the bathymetry has been lowered to that of the model's nearby open ocean: 580 m for the Indonesia experiment and 1,100 m for the deep-CAS experiment. The latter experiment shows the maximum possible signal of this particular tectonic change. Modified boundary conditions for the tectonic simulations can be provided on request. Variations of this experiment that have been performed previously were collated in ref. 41. Supplementary Fig. 12 shows how the SST changes in the present simulations compare with their compilation.

In the enhanced ocean mixing experiment, the hurricane-induced mixing is treated as an increase in background vertical diffusivity of 1 cm<sup>2</sup> s<sup>-1</sup> in two subtropical bands (8° N–40° N and 8° S–40° S) at depths down to the model level at 185 m (ref. 45; Supplementary Fig. 13b). The value of 1 cm<sup>2</sup> s<sup>-1</sup> was chosen as it is the annual mean diffusivity observed in regions of strong tropical cyclone activity<sup>81</sup>. A recent study shows that imposing permanent mixing in the upper ocean has a largely similar effect as a transient mixing<sup>82</sup>.

In the reduced-cloud-albedo experiment, the cloud liquid water path has been reduced by 20% polewards of 30° N and 30° S, but only for the purposes of calculating the shortwave radiation. This acts to reduce the albedo of highly reflective low stratus clouds (Supplementary Fig. 13a). As a result, global mean air surface temperature increased by 1.1 °C (Supplementary Table 4). One of the effects of this modification was an increase of 0.7 °C in the tropical West Pacific SST, which stopped us from imposing a greater reduction. We note that in this experiment we do not directly prescribe changes in cloud albedo, which is calculated within the CESM. This is a different approach from that used to investigate sensitivity to cloud properties in ref. 44. Those authors used a coupled model of intermediate complexity with a specified cloud cover, which they reduced by 50% in their perturbation experiments. Here we simulated a range of reduced liquid water paths in the atmosphere, with a 20% reduction leading to the representation closest to that of ref. 44. The actual change for the early Pliocene, if any, is largely unconstrained.

We note that the effects of changes in extratropical low clouds and enhanced ocean mixing on tropical climate seem to be robust in a variety of models. New experiments are now being conducted with high-resolution versions of CESM and other coupled models, which seem to confirm our previous conclusions (to be reported elsewhere).

Among hypotheses that we have not tested, we should mention atmospheric superrotation<sup>49</sup>. In such a state, common on gas giants such as Jupiter, atmospheric Rossby waves radiate polewards from the tropics, generating westerly winds along the Equator and suppressing equatorial upwelling. Recent studies<sup>83</sup>, however, indicate that superrotation develops only at temperatures in excess of 40 °C. It has also been argued that a stronger meridional salinity gradient (with more fresh water in the northern Pacific) could reduce oceanic meridional density gradient and collapse the tropical thermocline<sup>48,84</sup>. This would require a strong increase in precipitation in high latitudes, for which evidence has yet to be found.

**Carbon dioxide estimates.** The CO<sub>2</sub> estimates are calculated using a variety of techniques based on carbon isotopes<sup>46</sup>, boron isotopes<sup>85–87</sup>, alkenones<sup>38,87</sup>, boron/calcium ratios<sup>88</sup>, stomata<sup>89</sup> and ice-core inclusions<sup>90–92</sup> (Supplementary Table 1). They are presented as originally published except for the record based on  $\delta^{13}\text{C}$  (ref. 46), which was adjusted to subtract an offset (90 p.p.m.) between the original data and the Antarctic ice-core CO<sub>2</sub> records.

53. Lawrence, K. T., Herbert, T. D., Dekens, P. S. & Ravelo, A. in *Deep-Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies* (eds Williams, M., Haywood, A. M., Gregory, F. J. & Schmidt, D. N.) 539–562 (Geol. Soc., 2007).
54. Müller, P., Kirst, G., Ruhland, G., von Storch, I. & Rosell-Melé, A. Calibration of the alkenone paleotemperature index  $U^{K'}_{37}$  based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S). *Geochim. Cosmochim. Acta* **62**, 1757–1772 (1998).

55. Martínez-García, A., Rosell-Melé, A., McClymont, E. L., Gersonde, R. & Haug, G. H. Subpolar link to the emergence of the modern equatorial pacific cold tongue. *Science* **328**, 1550–1553 (2010).
56. McClymont, E. L., Rosell-Melé, A., Giraudieu, J., Pierre, C. & Lloyd, J. M. Alkenone and coccolith records of the mid-Pleistocene in the south-east Atlantic: implications for the index and south African climate. *Quat. Sci. Rev.* **24**, 1559–1572 (2005).
57. Müller, P. J., Cepek, M., Ruhland, G. & Schneider, R. R. Alkenone and coccolithophorid species changes in late quaternary sediments from the Walvis ridge: implications for the alkenone paleotemperature method. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **135**, 71–96 (1997).
58. Dekens, P., Lea, D., Pak, D. & Spero, H. Core top calibration of Mg/Ca in tropical foraminifera: refining paleotemperature estimation. *Geochem. Geophys. Geosyst.* **3**, 1–29 (2002).
59. Anand, P., Elderfield, H. & Conte, M. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18**, 1050 (2003).
60. Karas, C., Nürnberg, D., Tiedemann, R. & Garbe-Schönberg, D. Pliocene Indonesian throughflow and Leeuwin current dynamics: implications for Indian Ocean polar heat flux. *Paleoceanography* **26**, PA2217 (2011).
61. Regenberk, M. et al. Assessing the effect of dissolution on planktonic foraminiferal Mg/Ca ratios: evidence from Caribbean core tops. *Geochem. Geophys. Geosyst.* **7**, Q07P15 (2006).
62. Groeneveld, J. et al. Foraminiferal Mg/Ca increase in the Caribbean during the Pliocene: western Atlantic warm pool formation, salinity influence, or diagenetic overprint. *Geochem. Geophys. Geosyst.* **9**, Q01P23 (2008).
63. Groeneveld, J. *Effect of the Pliocene Closure of the Panamanian Gateway on Caribbean and east Pacific Sea Surface Temperatures and Salinities by Applying Combined Mg/Ca and  $^{18}\text{O}$  Measurements (5.6–2.2 Ma)* 9–14. PhD thesis, Univ. Kiel (2005).
64. de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F. & Beaufort, L. Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature* **433**, 294–298 (2005).
65. Karas, C. et al. Mid-Pliocene climate change amplified by a switch in Indonesian subsurface throughflow. *Nature Geosci.* **2**, 434–438 (2009).
66. Etourneau, J., Martinez, P., Blanz, T. & Schneider, R. Pliocene-Pleistocene variability of upwelling activity, productivity, and nutrient cycling in the Benguela region. *Geology* **37**, 871–874 (2009).
67. Dekens, P. S., Ravelo, A. C. & McCarthy, M. D. Warm upwelling regions in the Pliocene warm period. *Paleoceanography* **22**, PA3211 (2007).
68. Liu, Z. & Herbert, T. D. High-latitude influence on the eastern equatorial Pacific climate in the early Pleistocene epoch. *Nature* **427**, 720–723 (2004).
69. Herbert, T., Peterson, L., Lawrence, K. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328**, 1530–1534 (2010).
70. Cleaveland, L. C. & Herbert, T. D. Coherent obliquity band and heterogeneous precession band responses in early Pleistocene tropical sea surface temperatures. *Paleoceanography* **22**, PA2216 (2007).
71. Etourneau, J., Martinez, P., Blanz, T. & Schneider, R. Intensification of the Walker and Hadley atmospheric circulations during the Pliocene-Pleistocene climate transition. *Earth Planet. Sci. Lett.* **297**, 103–110 (2010).
72. LaRiviere, J. et al. Late Miocene decoupling of oceanic warmth and atmospheric carbon dioxide forcing. *Nature* **486**, 97–100 (2012).
73. Herbert, T. D. & Schuffert, J. Alkenone unsaturation estimates of late Miocene through late Pliocene sea surface temperature changes, ODP site 958. *Proc. ODP Sci. Res.* **159T**, 17–22 (1998).
74. Martínez-García, A. et al. Links between iron supply, marine productivity, sea surface temperature, and CO<sub>2</sub> over the last 1.1 Ma. *Paleoceanography* **24**, PA1207 (2009).
75. Lawrence, K., Sosdian, S., White, H. & Rosenthal, Y. North Atlantic climate evolution through the Plio-Pleistocene climate transitions. *Earth Planet. Sci. Lett.* **300**, 329–342 (2010).
76. Karas, C., Nürnberg, D., Tiedemann, R. & Garbe-Schönberg, D. Pliocene climate change of the southwest Pacific and the impact of ocean gateways. *Earth Planet. Sci. Lett.* **301**, 117–124 (2011).
77. Tian, J. et al. Late Pliocene monsoon linkage in the tropical South China Sea. *Earth Planet. Sci. Lett.* **252**, 72–81 (2006).
78. Ravelo, A. Walker circulation and global warming: lessons from the geologic past. *Oceanography* **19**, 114–122 (2006).
79. Shields, C. A. et al. The low resolution CCSM4. *J. Clim.* **25**, 3993–4014 (2012).
80. Bitz, C. M. et al. Climate sensitivity of the Community Climate System Model version 4. *J. Clim.* **25**, 3053–3070 (2012).
81. Srivastava, R. L. & Huber, M. Observational evidence for an ocean heat pump induced by tropical cyclones. *Nature* **447**, 577–580 (2007).
82. Manucharyan, G., Brierley, C. & Fedorov, A. Climate impacts of intermittent upper ocean mixing induced by tropical cyclones. *J. Geophys. Res.* **116**, C11038 (2011).
83. Caballero, R. & Huber, M. Spontaneous transition to superrotation in warm climates simulated by CAM3. *Geophys. Res. Lett.* **37**, L11701 (2010).
84. McKay, R. et al. Antarctic and Southern Ocean influences on late Pliocene global cooling. *Proc. Natl. Acad. Sci. USA* **109**, 6423–6428 (2012).
85. Bartoli, G., Höönicke, B. & Zeebe, R. E. Atmospheric CO<sub>2</sub> decline during the Pliocene intensification of Northern Hemisphere glaciations. *Paleoceanography* **26**, PA4213 (2011).
86. Höönicke, B., Hemming, N. G., Archer, D., Siddall, M. & McManus, J. F. Atmospheric carbon dioxide concentration across the mid-Pleistocene transition. *Science* **324**, 1551–1554 (2009).
87. Seki, O. et al. Alkenone and boron-based Pliocene pCO<sub>2</sub> records. *Earth Planet. Sci. Lett.* **292**, 201–211 (2010).
88. Tripati, A. K., Roberts, C. D. & Eagle, R. A. Coupling of CO<sub>2</sub> and ice sheet stability over major climate transitions of the last 20 million years. *Science* **326**, 1394–1397 (2009).
89. Kürschner, W. M., van der Burgh, J., Visscher, H. & Dilcher, D. L. Oak leaves as biosensors of late Neogene and early Pleistocene paleoatmospheric CO<sub>2</sub> concentrations. *Mar. Micropaleontol.* **27**, 299–312 (1996).
90. Petit, J. et al. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
91. Siegenthaler, U. et al. Stable carbon cycle-climate relationship during the late Pleistocene. *Science* **310**, 1313–1317 (2005).
92. Lüthi, D. et al. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).

# Intra-oceanic subduction shaped the assembly of Cordilleran North America

Karin Sigloch<sup>1</sup> & Mitchell G. Mihalynuk<sup>2</sup>

The western quarter of North America consists of accreted terranes—crustal blocks added over the past 200 million years—but the reason for this is unclear. The widely accepted explanation posits that the oceanic Farallon plate acted as a conveyor belt, sweeping terranes into the continental margin while subducting under it. Here we show that this hypothesis, which fails to explain many terrane complexities, is also inconsistent with new tomographic images of lower-mantle slabs, and with their locations relative to plate reconstructions. We offer a reinterpretation of North American palaeogeography and test it quantitatively: collision events are clearly recorded by slab geometry, and can be time calibrated and reconciled with plate reconstructions and surface geology. The seas west of Cretaceous North America must have resembled today's western Pacific, strung with island arcs. All proto-Pacific plates initially subducted into almost stationary, intra-oceanic trenches, and accumulated below as massive vertical slab walls. Above the slabs, long-lived volcanic archipelagos and subduction complexes grew. Crustal accretion occurred when North America overrode the archipelagos, causing major episodes of Cordilleran mountain building.

Continents grow through subduction magmatism and collision of arcs and other buoyant crustal fragments at their margins. Poorly understood, such collisions are of broad scientific interest because they cause rapid geographic changes, affecting climate, ocean circulation, biota and the formation of economically important metal deposits. North America was enlarged by a sequence of massive terrane collisions relatively recently (between 200 million years (Myr) ago and 50 Myr ago), which created the mountainous Cordillera of the American West<sup>1</sup>.

Reconciling geological records on land with those of the ocean basins has proved difficult. Magnetic stripes on the sea floor are the basis of all quantitative plate tectonic reconstructions, and well-preserved Atlantic spreading records indicate that North America has moved westward continuously since the breakup of Pangaea (about 185 Myr ago), away from Africa and Europe<sup>2</sup>. In contrast, more than half of the seafloor records of proto-Pacific (Panthalassa) ocean spreading are missing. The Pacific plate records the existence of another major oceanic plate to its northeast since at least 180 Myr ago: the Farallon plate. This plate is usually assumed to have filled the eastern Panthalassa basin, extending to the western margin of North America and subducting under it, although this is not required by the magnetic seafloor data. Hence, the Farallon plate, invoked as the causative agent in almost all major land geological events since late Jurassic times<sup>2,3</sup>, should also have transported the terranes to the continental margin.

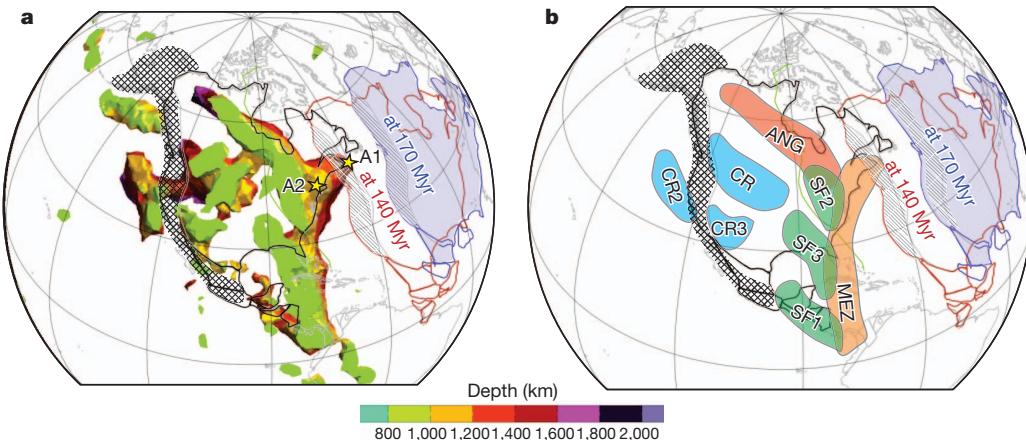
However, dozens of terranes have accreted to North America since 200 Myr ago<sup>1</sup>, but not to the Andean margin of South America, supposedly a closely analogous setting. The terranes are mostly Triassic to Cretaceous island arc–subduction assemblages, but include three more heterogeneous superterrane: Intermontane (IMS)<sup>4,5</sup>, Insular (INS)<sup>4</sup> and Guerrero (GUE)<sup>6</sup>, which are essentially microcontinents. Their exact origins remain mysterious, but the terranes formed at various latitudes and times, as inferred from palaeomagnetic observations and fossil faunas, and pre-assembled at others. This implies the temporary existence of additional oceanic plates in the northeastern proto-Pacific Ocean<sup>7,8</sup>, which are missing from quantitative plate reconstructions.

The mantle retains a memory of ancient plate configurations, in the form of subducted slabs, which body-wave tomography images *in situ* as domains of faster-than-average seismic wave velocities. Beneath North America these slab relics are massive, almost vertical walls extending from 800 to 2,000 km in depth, and typically 400–600 km wide (Fig. 1, Supplementary Fig. 1). The largest wall runs from northwest Canada to the eastern USA and on to Central America, and has been called the Farallon slab, one of the most massive features in global tomographies<sup>9–12</sup>. We argue that in fact most of this slab wall is not Farallon and subdivide it into the Angayucham (ANG), Mezcalera (MEZ) and the Southern Farallon (SF) components (Fig. 1b). This reinterpretation is based on our most recent tomography model<sup>13</sup>, which utilizes dense USArray data<sup>14</sup> in addition to global network data, using a cutting-edge waveform inversion method: multiple-frequency P-wave tomography<sup>13,15</sup>.

Besides putting the known eastern slab walls<sup>9–12</sup> in sharper focus, we discovered<sup>15</sup> that another, more westerly slab wall, the Cascadia Root (CR in Fig. 1b, depth 700 km to more than 1,800 km), connects upward continuously to the present-day Cascadia trench, into which the last remnant of the Farallon (Juan de Fuca) plate is subducting. This makes CR a Farallon slab by definition, and prompted us to re-evaluate whether MEZ/ANG/SF do really constitute subducted Farallon Ocean floor, which to our knowledge has never been questioned.

For plate reconstructions, the crucial question is whether (and how) these lower-mantle slab walls have moved laterally since they were deposited beneath their corresponding volcanic arcs (past versus present *x*-*y* positions in an absolute reference frame). Here we quantitatively test the hypothesis that they have not moved appreciably, that is, there has been only vertical sinking within our observational uncertainties (a few hundred kilometres laterally). Motivated primarily by imaged slab geometries, this null hypothesis also seems sensible in light of the Cenozoic subduction record, where absolute trench motion on average contributed only 10–30% of total plate convergence<sup>16,17</sup>.

<sup>1</sup>Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität, Theresienstrasse 41, 80333 Munich, Germany. <sup>2</sup>British Columbia Geological Survey, PO Box 9333 Stn Prov Govt, Victoria, British Columbia V8W 9N3, Canada.



**Figure 1 | Slabs under North America and continental motion over time.** **a**, Subducted slabs at and below 900 km depth. P-wave tomography model<sup>13</sup> rendered as three-dimensional (3D) isosurface contours, which enclose faster-than-average structure (threshold  $d\nu_p/\nu_p = 0.25\%$ , where  $\nu_p$  is the P-wave velocity). Colour signifies depth and changes every 200 km; the scene is illuminated to convey 3D perspective. At a sinking rate of about  $10 \text{ mm yr}^{-1}$ , this slab assemblage should have been deposited from about 200 Myr ago to 90 Myr ago. Reconstructed continent positions at 140 Myr ago are shown in a hotspot reference frame<sup>21</sup> and at 170 Myr ago in a hotspot/palaeomagnetic

hybrid frame<sup>22</sup>. The hatched area represents location uncertainty for continental margin during Jurassic/Cretaceous times; the cross-hatched area shows terranes that accreted during Cretaceous and early Tertiary times. **b**, Interpretive legend. The slab walls divide into four groups: Cascadia/Northern Farallon slabs (blue) and Southern Farallon slabs (green), owing to eastward subduction; Angayucham (ANG, red) and Mezcalera (MEZ, orange) slabs, owing to westward subduction. Before 140 Myr ago, sizeable ocean basins separated North America from the ANG/MEZ trenches.

### Slabs and arcs at stationary trenches

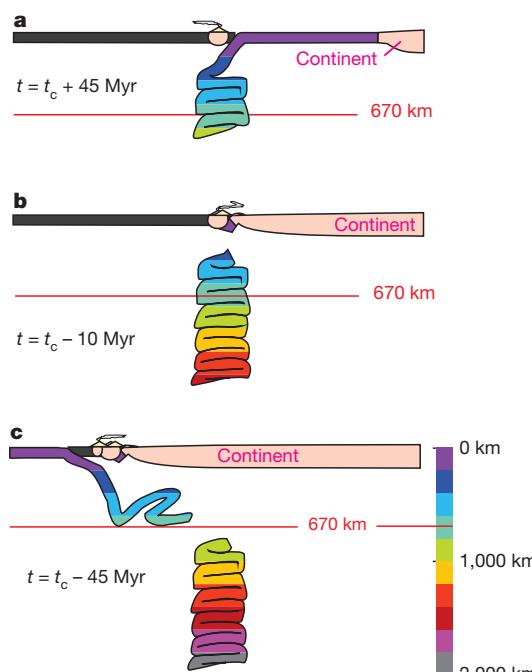
Figure 2a and b shows how a steep, widened slab wall could be piled up by nearly vertical sinking beneath a long-lived, stationary trench and volcanic arc. An Andean-style west-coast trench could not have been stationary because North America moved westward as the Atlantic Ocean spread. This contradiction is resolved by westward intra-oceanic

subduction before the arrival of the continent, followed by a polarity switch of subduction to its current eastward motion into a continental-margin trench (Fig. 2c). Such a scenario implies that the imaged lower-mantle slabs MEZ/ANG/SF are Jurassic to Cretaceous in age, allowing the collision of North America with their subduction zones to cause the Cretaceous terrane accretions.

To the extent that slabs sink vertically, they record palaeo-arc and trench locations in an absolute sense. Thus, vertical sinking permits quantitative predictions of the location and timing of continent-trench collisions when tomography and absolute plate reconstructions are combined. These predictions can be tested against the docking times of arc terranes inferred from land geology. Abrupt upward truncations of the slab walls, which are well resolved tomographically (Supplementary Fig. 2), correspond to the shutdown of the overlying trench-arc systems, and hence to docking times (Fig. 2).

If the trench remains stationary, a vertical slab pile is deposited beneath it. If the trench moves (but every parcel of slab sinks vertically), the imaged slab will dip towards older trench locations, assuming no dramatic lateral variations in sinking rate. The observed lower-mantle walls are widened to 400–600 km laterally, that is, 4–6 times the thickness of oceanic lithosphere—this is not artificial blur, but the actual reason for their robust tomographic visibility<sup>18</sup>. Thickening is probably achieved by slab folding above the 670-km viscosity jump, deviations from vertical sinking being due mainly to the folding process itself (Fig. 2). In convection models, slab folding occurs preferentially beneath the kind of stationary trenches postulated here<sup>18–20</sup>. Massive, thickened slabs like these can be expected to be the drivers of ‘mantle wind’, rather than blowing in it: that is, if anything sinks vertically, it should be these slabs.

Such slab walls indicate that their overlying trenches remained in the same absolute locations for a long time, with arc and accretionary complex growth stationed above these locations. Observation of massive slab walls leads us to think of their associated, intra-oceanic trenches as ‘terrane stations’ where new crustal material is gathered to await transfer to a continental margin. Terrane stations above ANG and MEZ were not conveyed eastward into a continental Farallon trench. Rather, North America migrated westward, collided, and accreted the ANG and MEZ terrane stations. Hence slab walls tie the now-displaced terranes to a laterally unchanged subsurface, constraining absolute locations and temporal evolution of oceanic trenches more than a hundred million years after their demise.



**Figure 2 | Schematic cross-section and evolution of a terrane station.**  $t_c$  denotes the time of arc-continent collision. Motions are shown in a lower-mantle reference frame. **a**, Well before the collision, both trench and arc are active. Slab buckling is due to the viscosity contrast around 670 km, but the backlog reaches into the upper mantle. **b**, Around  $t_c$  and up to about 10 Myr later, the continent overrides the trench and accretes its arc terranes, while the slab breaks. **c**, Well after the collision, the slab wall continues to sink. Seaward, a new Andean-style subduction has developed. Anchored in the lower mantle, the slab wall is sinking vertically at a steady-state rate of approximately  $10 \text{ mm yr}^{-1}$  in all three panels.

Figure 1 shows the reconstructed western margin of North America<sup>21</sup>, together with schematic outlines of the lower-mantle slab walls from tomography<sup>13</sup>. North America's relative westward motion is well constrained by the Atlantic spreading record<sup>2,22</sup>, independently of any absolute reference frame. Comparison to the basic geometry of ANG and MEZ suggests that these two slabs did not subduct beneath the continental margin: (1) the slabs are vertical from 800 to 2,000 km depth, indicating a stationary trench, whereas the margin moved westward continuously. (2) The outlines of ANG and MEZ, especially the pronounced eastward promontory of MEZ, do not match the outlines of the continental margin (Fig. 1). If continental subduction had controlled slab deposition, then slab curvature should reflect the curvature of the continent. (3) West of ANG/MEZ, the slab is smeared out laterally in the upper 800 km (Supplementary Fig. 1), as might be expected from a trench dragged along by a migrating continent: direct observational evidence for a switch in subduction mode after override, from stationary oceanic to migrating continental.

### Quantitative prediction of arc accretion

Continent motions in Figs 1 and 3 are tied to an absolute hotspot reference frame<sup>21</sup> and rendered by the palaeo-geographic information system GPlates<sup>23,24</sup>. Like the vertically sinking slab walls, vertically rising plumes are thought not to have significant lateral motions relative to the lowermost mantle<sup>25</sup> (smaller deviations are correctable<sup>21</sup>), so that the hotspot and 'slab wall' reference frames are equivalent (no relative motion). In this merged reference frame, the lateral overlay of North America's reconstructed western margin with a slab wall amounts to a spatiotemporal prediction of trench override and terrane accretion.

For example, Fig. 1a shows the North American margin overriding point A1, eastern promontory of the MEZ arc, sometime before 140 Myr, probably around 150 Myr ago. Slab sinking velocities can be inferred: the shallow end of MEZ beneath A1 is seen to have sunk to a depth of approximately 1,500 km, implying an averaged sinking rate of  $10 \text{ mm yr}^{-1}$ . The MEZ promontory shallows to the southwest. This is unrelated to trench polarity (the slab did not dip northeastward), but rather reflects differential sinking times (subduction at A1 was choked off earlier than at A2). Sinking rates could be estimated from any well-resolved point on the upward truncation of a slab wall, but we choose five points A1–A5 that are associated with supporting evidence from land geology (Table 1 and Fig. 3). Predicted override ages are Jurassic–Cretaceous (146–55 Myr ago), becoming progressively younger westward, and truncation depths shallow to the west as expected. Sinking rate estimates range between 9 and  $12 \text{ mm yr}^{-1}$  ( $\pm 2 \text{ mm yr}^{-1}$ ), consistent with findings of  $12 \pm 3 \text{ mm yr}^{-1}$  globally<sup>12</sup>. Figure 3 renders the override sequence in four time slices, each showing only slabs that should already have been deposited at the time, assuming the sinking rate was  $10 \text{ mm yr}^{-1}$  (the average of A1–A5 in Table 1).

The CR must be a Farallon slab because the Farallon (Juan de Fuca) plate is still subducting into it today<sup>15</sup>. Pacific seafloor records indicate continuous Farallon spreading since about 180 Myr ago<sup>3,26</sup>, so that at a sinking rate of  $10 \text{ mm yr}^{-1}$ , the over-1,800-km-deep CR accounts for the entire lifetime of the (northern) Farallon plate. The presence of this CR slab implies that the equally deep and thickened ANG slab further east cannot represent Farallon lithosphere, as has been assumed<sup>9,12,13,27,28</sup>. Rather, the ANG slab must have dipped in the opposite direction (southwestward) in order to have sourced sufficient plate material from an ocean basin that lay to the northeast, the consumption of which accommodated the westward drift of North America. This scenario for transporting North America away from the former Pangaea provides an alternative to westward rollback of a continental Farallon trench.

Hence our inferred trench/plate evolution in Fig. 3 differs fundamentally from the commonly accepted scenario of MEZ/ANG as products of east-dipping, Farallon-beneath-continent subduction.

Westward subduction of the ANG and MEZ slabs, probably since early Jurassic times, consumed the ocean that bounded North America on the west (the ANG and MEZ basins in Fig. 3). Both basins were consumed in a zipper-like fashion: MEZ closed from north to south, ANG from south to north, as North America gradually overrode and shut down the ANG/MEZ arcs between 150 and 50 Myr ago. Further west, the early Farallon Ocean subducted into two east–west-oriented Cascadia slabs (CR/CR2), but established additional segments SF1/SF2 after a clockwise rotation at about 147 Myr ago<sup>22,26</sup>. Thus at northerly latitudes, two long-lived terrane stations of opposite polarity coexisted in ANG and CR, a variant not considered in Fig. 2. Upon override, the east-dipping Farallon trenches started rolling back with North America. Moderate complexities in the Pacific–Farallon spreading record<sup>3,26</sup> probably reflect the transitions of individual trench segments from intra-oceanic to Andean-style.

### Supporting evidence from land geology

We now use the terrane-station property of oceanic trench/arc systems to test archipelago override predictions made by tomography and plate reconstructions. The collision of North America with buoyant arc crust should coincide with observed deformation and accretion events.

Figure 3b shows inferred terrane locations before override: each active trench/arc system may include a subduction complex or exotic fragments. Using geological relationships in the present-day Cordillera<sup>1</sup>, we can match most hypothesized terranes with actual ones. ANG terranes (red) now make up the interior of Alaska, in fault contact with the Angayucham and related ophiolites. Those studying Alaska have long inferred a southwest-dipping subduction<sup>29</sup>.

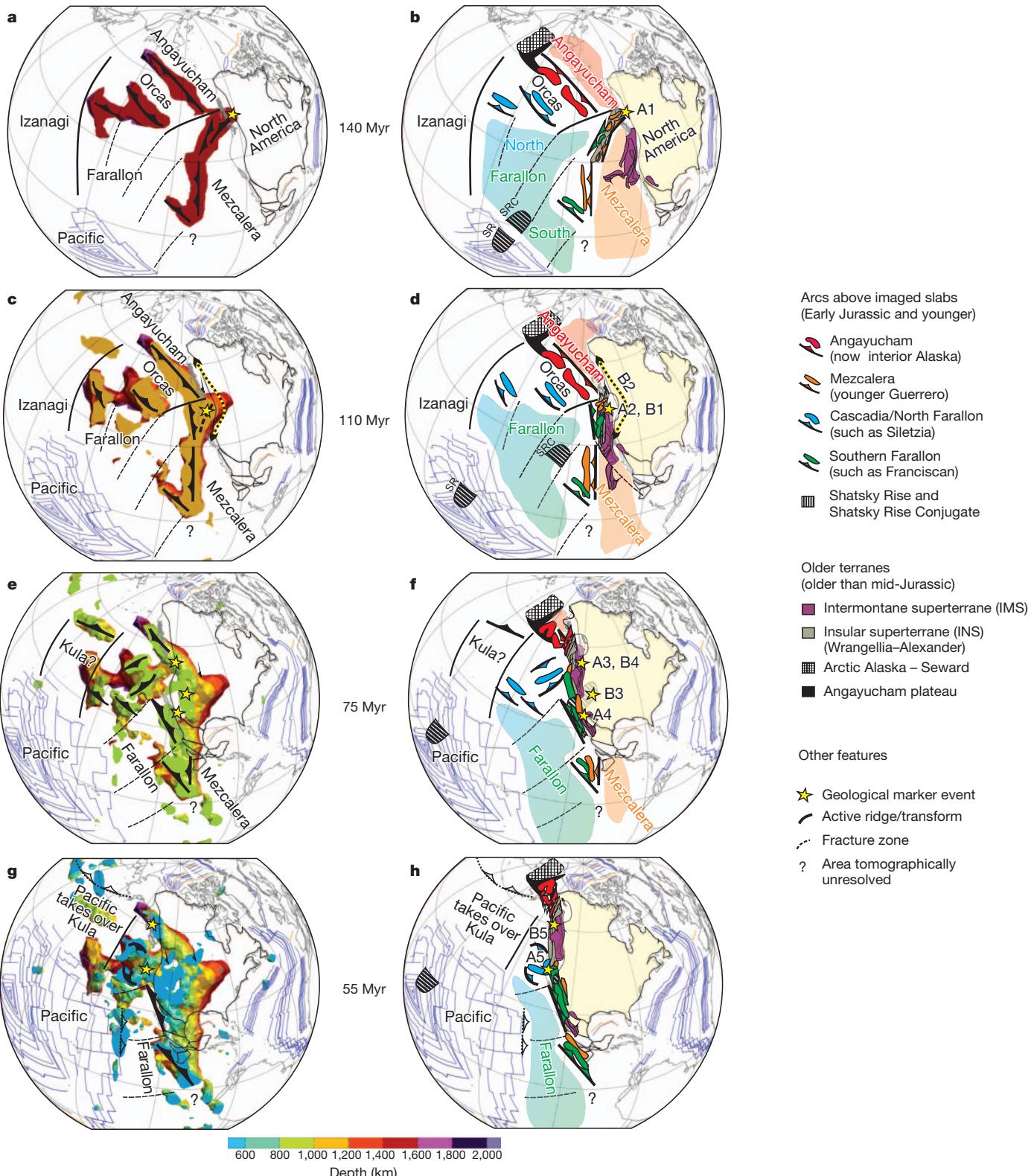
Green terranes west of A1 represent the Franciscan subduction complex of present-day California. Two superterrane from earlier subduction had already loosely accreted to North America before archipelago override began: the IMS closest to the continent<sup>4,5</sup>, and the GUE to the south<sup>6</sup>, whereas the INS superterrane<sup>4</sup> probably provided the subduction nucleation for the MEZ arc.

To provide for independent validation, our calibration points for sinking rates were chosen at tectonic events that are sharply defined in time and space (A1–A5 in Table 1). Events B1–B5, which are interleaved with the A1–A5 events, represent widespread Cordilleran orogenic and accretion episodes, demonstrating explanatory power on a continental scale. Four stages of override are distinguished, as follows.

Stage 1 (see Fig. 3a, b) is the beginning of the override of the east-verging MEZ promontory. Deformation was initially localized to the Pacific Northwest, as predicted by our model. Incipient deformation of the hinterland generated molasse that flooded the continental platform about 157 Myr ago between 45 and 55°N (ref. 30). A flip in subduction direction at about 165 Myr ago<sup>31</sup> is recorded by the transition from proto-Franciscan formation (for example, the Red Ant formation<sup>32</sup>, shown as orange terranes southwest of A1 in Fig. 3b) to Franciscan formation (green), marking an early subduction hand-over from MEZ to SF2.

Stage 2 (Figs 3c, d) is the time of margin-wide orogenies as North America collided with an increasingly wide swath of MEZ/SF. This caused the Sevier and Canadian Rocky Mountains orogenesis since around 125 Myr ago. Inboard parts of IMS were partly constructed on top of stable North American crust in southern California<sup>33</sup> and had largely collapsed by 110 Myr ago, increasingly shedding zircons onto stable North America<sup>34</sup>, and vice versa; whereas IMS and the active Sierra Nevada arc shed zircons into the Franciscan trench<sup>33</sup> (SF2). Intrusion of Omenica magmatic belts successively eastward into northern IMS and adjacent displaced North American strata<sup>8</sup> (B1, about 124–90 Myr ago) can be attributed to prolonged override of the MEZ promontory.

Stage 3 (Fig. 3e, f) is when North America entered the Farallon hemisphere. As ANG collided obliquely, its terranes (red; now interior Alaska) were accreted along the Canadian margin. Override of A3



**Figure 3 | Sequence of trench overrides and terrane accretions.** The left column (a, c, e, g) shows time-depth slices at  $t = 140, 110, 75$  and  $55$  Myr ago; the tomography model and plate reconstructions are rendered as in Fig. 1. Each slice shows only material that should have been deposited by that time, that is, slab at and below a depth of  $v \times t$ , where  $v = 10 \text{ mm yr}^{-1}$  is the assumed sinking rate. All slabs are coloured according to their current depths, but mentally their upper truncations should be migrated up to the surface, representing the inferred active arc locations at each time. The 140-Myr-ago

slice renders slab below 1,400 km depth; the 110-Myr-ago slice renders slab below 1,100 km; the 75-Myr-ago slice below 750 km; and the 55-Myr-ago slice below 550 km. Blue lines are preserved seafloor isochrons. The associated maps in the right column (b, d, f, h) are interpretative cartoons showing the evolution of inferred trench and terrane geometries. Yellow stars mark the tectonic events of Table 1. SR, Shatsky Rise; SRC, Shatsky Rise Conjugate.

**Table 1 | Sequence of archipelago override**

| Event | Geometric/kinematic event description   | Matched geological validation event  | Reconstructed time,<br><i>t</i> (Myr ago) | Slab depth,<br><i>d</i> (km) | Slab sinking velocity,<br><i>v</i> (mm yr <sup>-1</sup> ) |
|-------|---|--|---|------------------------------|---|
| A1    | Start override of MEZ promontory. Overridden segment is replaced by incipient South Farallon trench SF2.    | Initiation of Rocky Mountain deformation, recorded by synorogenic clastic wedge (160–155 Myr ago). Initiation of Franciscan subduction complex/South Farallon (165–155 Myr ago). | 146 ± 24                                  | 1,500 ± 100                  | 10 ± 2  |
| B1    | Gradual override of MEZ promontory by North America (Pacific Northwest).                                    | Omenica magmatic belts in Pacific Northwest (124–90 Myr ago).  | -   | -                            | -   |
| A2    | End override of MEZ promontory (shallowest point of slab wall).   | -  | 111 ± 8                                   | 1,050 ± 50                   | 9 ± 1   |
| B2    | Widening collision of North America margin with archipelago (MEZ/ANG/SF arcs).                              | Margin-wide strong deformation: Sevier and Canadian Rocky Mountains (since about 125 Myr ago).   | -   | -                            | -   |
| A3    | Override of ANG arc, followed by slab window.   | Carmacks volcanic episode due to slab window (72–69 Myr ago).  | 74 ± 7                                    | 850 ± 50                     | 12 ± 1  |
| A4    | South Farallon trench steps westward after accretion of Shatsky Rise Conjugate plateau.                     | Sonora volcanism due to slab window: Tarahumara ignimbrite province (85 ± 5 Myr ago).  | 88 ± 3                                    | 800 ± 50                     | 9 ± 1   |
| B3    | Strong transpressive coupling of Farallon plate to superterrane as buoyant Shatsky Rise Conjugate subducts. | Laramide orogeny, basement uplift more than 1,000 km inland (85–55 Myr ago).   | -   | -                            | -   |
| B4    |   | Northward shuffle of INS, IMS, and ANG terranes along margin (85–55 Myr ago).  | -   | -                            | -   |
| A5    | Override of CR arc by Pacific Northwest USA   | Last terrane accretions: Siletzia, Pacific Rim (55–50 Myr ago).  | 55 ± 7                                    | 600 ± 30                     | 12 ± 2  |
| B5    | Final override of westernmost ANG.  | Explosive end of Coast Mountain arc volcanism (55–50 Myr ago).   | -   | -                            | -   |

Column 2 describes tectonic events in terms of geometrically predicted slab–margin interactions. Column 3 describes matching events from the land geological record. For well-localized events (A1–A5), we estimate slab depth *d*, time *t* since last subduction (margin override), and slab sinking rate *v*, from the tomographic and plate models. Calculations are explained in the Methods; for uncertainty analysis see the Supplementary Information. Geologically observed timing does not enter the calculations, since the role of the geological events is to validate the geometrically inferred results. Events B1–B5 represent additional, first-order tectonic episodes explained by the scenario of archipelago override (these are not localized enough spatiotemporally to estimate *d*, *t* and *v*).

was accompanied by a strong pulse of intermediate to basaltic volcanism, the Carmacks formation at around 72–69 Myr ago<sup>35</sup>. Such high-temperature, mainly primitive volcanism arises from juxtaposition of hot, sub-slab asthenosphere after a slab, broken through arc collision (Fig. 2b), has started to sink<sup>36</sup>.

The South Farallon trench was migrating westward with North America, still building the Franciscan formation. Around A4, slab geometry indicates an outboard step from SF2 to SF3. This coincides with a strong regional pulse of ignimbrite volcanism at 90 Myr ago (Tarahumara formation, up to 4 km thick and 400 km inland<sup>37,38</sup>) as the southern California/Sonora margin traverses the intermittent slab gap.

The location and timing of this trench step-back coincide almost perfectly with the inferred arrival of a buoyant oceanic Farallon plateau at the North America margin, the conjugate half of the Shatsky Rise (Fig. 3 includes its reconstruction by ref. 39). Plateau collision is a proposed mechanism<sup>39,40</sup> for choking subduction and causing the basement uplifts of the Laramide orogeny at around 85–55 Myr ago (B3). We suggest that the event explains another first-order observation of Cordilleran geology (B4): the INS/IMS and ANG terrane packages, but not Franciscan and GUE, were rapidly shuffled northward along the margin between 85 and 55 Myr ago, by many hundreds to over 2,000 km (ref. 41). The convergence vector of the Farallon plate did have a large northward component at the time<sup>2</sup>, but terrane transport additionally requires strong coupling to the Farallon plate and decoupling from North America. A buoyant Farallon plateau, compressed against INS/IMS and unable to subduct, could have achieved such coupling until the trench re-established itself further west.

Stage 4 (Fig. 3g, h) is the end of archipelago override. North America overrides point A5 at 55 ± 7 Myr ago, in excellent agreement with last observed terrane accretions in the Pacific Northwest (Siletzia, Metchosin and Pacific Rim terranes at about 55–50 Myr ago<sup>42–44</sup>). The trench stepped west (clear upward truncation at A5) as the terranes accreted, converting intra-oceanic CR into today's continental Cascadia subduction. Also at about 55–50 Myr ago, ANG was terminally overridden (B5), accompanied by explosive volcanism as the Coast Mountain arc of British Columbia shut down.

By then, slab complexity in the upper mantle rivalled today's western Pacific—not surprising given the numerous forced reorganizations. The simple depth–age relationship suggested by Fig. 3 need not

apply to flat-lying Cenozoic slabs<sup>13</sup> or to isolated deeper fragments like Kula or CR3, because all sinking estimates were calibrated on slab walls. We do not attempt to interpret upper-mantle structure and the Cenozoic land record here, but our archipelago model provides a new framework for doing so.

## Better constraints on surface and mantle

Long-lived, stationary oceanic trenches explain two problematic observations, previously thought to be unconnected: the near-vertical geometries of the super-slabs under North America (without invoking exceptional mantle rheologies or *ad hoc* shifts in absolute reference frame), and the long series of arc terrane accretions during Cretaceous times<sup>1</sup> (not explained by the South American analogy).

An archipelago offshore Mesozoic North America had previously been suggested on the basis of land geology<sup>7,8</sup>, but lacked the absolute spatial constraints provided by seismic tomography and our vertical sinking/terrace stations concept. Now-displaced terranes can be tied to their original, seismically imaged trench locations, but also to reconstructed continent locations, in an absolute reference frame. The hypothesis that continental collision with stationary terranes caused the various episodes of Cordilleran mountain building becomes testable.

Before 150 Myr ago, North America was clearly located too far east for MEZ/ANG to pass for continental trenches, even when longitudinal uncertainties of absolute reference frames<sup>45</sup> are factored in (Fig. 1 and Supplementary Fig. 3, and Supplementary Tables 1 and 2). The apparently ‘wrong’ geometry and locations of the vertical slab walls—under the hypothesis of subduction of the Farallon plate beneath North America—had been recognized<sup>12,27,46,47</sup> with two kinds of solutions suggested. A longitudinal shift of the global lithospheric shell relative to the lower mantle<sup>12</sup>, specifically a Cretaceous westward excursion that tapered down as the Atlantic opened, could have held North America stationary above MEZ/ANG. Alternatively, upper-mantle slab, spread out laterally by a west-coast trench, must somehow have aggregated into steep piles when transitioning into the lower mantle. This requires lateral displacements by over 1,000 km of huge volumes of slab<sup>27,46</sup>. Some convection simulations have produced such behaviour<sup>46,48</sup>, whereas others suggest essentially vertical sinking<sup>47</sup>, as do newer observations<sup>12,13,18</sup>.

We showed that, within observational uncertainty limits, predicted and observed geological events are consistent. This validates simple vertical sinking and seems to explain all North American observations,

including accreted terranes, but is incompatible with the widely accepted continental Farallon trench since 175 Myr ago or before. The observed proportionality between slab depth and time since override (consistent sinking rates across all three slab walls) is not required for our argument, but rather increases confidence in its correctness.

Vertical slab sinking provides much tighter constraints on palaeogeographic reconstructions than arbitrarily movable slabs. This scenario follows the principle of parsimony, so that future investigations should start with it, and seek observations requiring a departure from it. The equivalence of hotspot and “slab wall” absolute reference frames, which is implied by vertical sinking, is of great interest because slabs reach back farther in time than hotspot tracks (200 Myr ago or more<sup>12,49</sup> versus about 130 Myr ago<sup>21,22</sup>), and they constrain absolute palaeo-longitude, which palaeomagnetic data alone cannot<sup>50</sup>. However, to quantitatively realize a global subduction reference frame<sup>12</sup>, it will be necessary to re-examine whether trenches commonly assumed to have been continental were not actually intra-oceanic.

## METHODS SUMMARY

We postulate that subduction into all slab walls imaged tomographically beneath North America<sup>13</sup> originated before the arrival of North America’s western margin. Hence we must demonstrate sufficiently old slab ages, and cessation or flipping of subduction when the continent overrode the slabs. Plate reconstructions predict the timing of margin arrival above a slab, but only if slab<sup>13</sup> and plate reconstructions<sup>12,21,22,50–52</sup> can be linked to the same absolute reference frame. Hence override predictions are correct to the extent that slab walls sank vertically, meaning that their  $x-y$  locations since subduction are unchanged in a linked hotspot reference frame. Uncertainty is best quantified from slab wall geometry itself: deviation from vertical sinking probably did not exceed a wall’s half-width (200–300 km), else such steep geometries could not have built up over a long time.

Uncertainties about absolute locations of North America’s palaeo-margin arise from imperfections in plate reconstructions. Owing to terrane accretions and removals, there is also uncertainty about the shape and westward extent of North America’s margin, compared to its present-day outlines. We discuss individual uncertainties in the Supplementary Information, and propagate them into cumulative uncertainties for the times at which North America’s palaeo-margin overrode selected points A1–A5 on the palaeo-trenches. With relative uncertainties of only 10–15% (Table 1 and Supplementary Tables 1 and 2), this yields the old (Jura-Cretaceous) slab ages required to support intra-oceanic subduction.

Spatiotemporal predictions of trench override are verified by terrane observations: override should coincide with observable collision events, because buoyant island arcs or plateaus are overridden. Uncertainties on terrane observations are difficult to quantify, but particularly characteristic events (A1–A5) can nonetheless be singled out and used successfully for validation (Table 1). Clear upward truncations of all slab walls offer direct observational evidence for continental override of oceanic trenches, and are used to calculate slab-wall sinking rates.

**Full Methods** and any associated references are available in the online version of the paper.

Received 29 August 2012; accepted 14 February 2013.

- Coney, P. J., Jones, D. L. & Monger, J. W. H. Cordilleran suspect terranes. *Nature* **288**, 329–333 (1980).
- Engebretson, D. C., Cox, A. & Gordon, R. Relative motions between oceanic and continental plates in the Pacific Basin. *Geol. Soc. Am. Spec. Pap.* **206**, 1–59 (1985).
- Atwater, T. Plate Tectonic History of the Northeast Pacific and Western North America. In *The Eastern Pacific Ocean and Hawaii* (eds Winterer, E. L., Hussong, D. M. & Decker, R. W.) N, 21–71 (Geological Society of America, 1989).
- Monger, J. W. H., Price, R. A. & Tempelman-Kluit, D. J. Tectonic accretion and the origin of the two major metamorphic and plutonic belts in the Canadian Cordillera. *Geology* **10**, 70–75 (1982).
- Mihalynuk, M. G., Nelson, J. & Diakow, L. J. Cache Creek Terrane entrapment: oroclinal paradox within the Canadian Cordillera. *Tectonics* **13**, 575–595 (1994).
- Dickinson, W. R. & Lawton, T. F. Carboniferous to Cretaceous assembly and fragmentation of Mexico. *Geol. Soc. Am. Bull.* **113**, 1142–1160 (2001).
- Moores, E. M. Ophiolites, the Sierra Nevada, ‘Cordilleran,’ and orogeny along the Pacific and Caribbean margins of North and South America. *Int. Geol. Rev.* **40**, 40–54 (1998).
- Johnston, S. T. The great Alaskan terrane wreck; reconciliation of paleomagnetic and geological data in the northern Cordillera. *Earth Planet. Sci. Lett.* **193**, 259–272 (2001).
- Grand, S. P., Van der Hilst, R. D. & Widjiantoro, S. Global seismic tomography; a snapshot of convection in the Earth. *GSA Today* **7**, 1–7 (1997).
- Montelli, R., Nolet, G., Masters, G., Dahlen, F. A. & Hung, S.-H. Global P and PP traveltimes tomography: rays versus waves. *Geophys. J. Int.* **158**, 637–654 (2004).
- Li, C., Van der Hilst, R. D., Engdahl, E. R. & Burdick, S. A new global model for P wave speed variations in Earth’s mantle. *Geochem. Geophys. Geosyst.* **9**, Q05018, <http://dx.doi.org/10.1029/2007GC001806> (2008).
- van der Meer, D. G., Spakman, W., Van Hinsbergen, D. J. J., Amaru, M. L. & Torsvik, T. H. Towards absolute plate motions constrained by lower-mantle slab remnants. *Nature Geosci.* **3**, 36–40 (2010).
- Sigloch, K. Mantle provinces under North America from multifrequency P wave tomography. *Geochem. Geophys. Geosyst.* **12**, Q02W08, <http://dx.doi.org/10.1029/2010GC003421> (2011).
- Pavlis, G. L., Sigloch, K., Burdick, S., Fouch, M. J. & Vernon, F. L. Unraveling the geometry of the Farallon plate: synthesis of three-dimensional imaging results from USArray. *Tectonophysics* **532–535**, 82–102 (2012).
- Sigloch, K., McQuarrie, N. & Nolet, G. Two-stage subduction history under North America inferred from multiple-frequency tomography. *Nature Geosci.* **1**, 458–462 (2008).
- Sdrolias, M. & Müller, R. D. Controls on back-arc basin formation. *Geochem. Geophys. Geosyst.* **7**, Q04016, <http://dx.doi.org/10.1029/2005GC001090> (2006).
- Goes, S., Capitanio, F. A., Morra, G., Seton, M. & Giardini, D. Signatures of downgoing plate-buoyancy driven subduction in Cenozoic plate motions. *Phys. Earth Planet. Inter.* **184**, 1–13 (2011).
- Ribe, N. M., Stutzmann, E., Ren, Y. & Van der Hilst, R. Buckling instabilities of subducted lithosphere beneath the transition zone. *Earth Planet. Sci. Lett.* **254**, 173–179 (2007).
- Gibert, G., Gerbault, M., Hassani, R. & Tric, E. Dependency of slab geometry on absolute velocities and conditions for cyclicity: insights from numerical modelling. *Geophys. J. Int.* **189**, 747–760 (2012).
- Goes, S., Capitanio, F. A. & Morra, G. Evidence of lower-mantle slab penetration phases in plate motions. *Nature* **451**, 981–984 (2008).
- O’Neill, C., Müller, D. & Steinberger, B. On the uncertainties in hot spot reconstructions and the significance of moving hot spot reference frames. *Geochem. Geophys. Geosyst.* **6**, Q04003, <http://dx.doi.org/10.1029/2004GC000784> (2005).
- Seton, M. et al. Global continental and ocean basin reconstructions since 200 Ma. *Earth Sci. Rev.* **113**, 212–270 (2012).
- Boyden, J. A. et al. in *Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences* (eds Keller, G. R. & Baru, C.) Ch. 7, 95–114 (Cambridge Univ. Press, 2011).
- Gurnis, M. et al. Plate tectonic reconstructions with continuously closing plates. *Comput. Geosci.* **38**, 35–42 (2012).
- Morgan, W. J. Convection plumes in the lower mantle. *Nature* **230**, 42–43 (1971).
- Sager, W. W., Handschumacher, D. W., Hilde, T. W. C. & Bracey, D. R. Tectonic evolution of the northern Pacific plate and Pacific-Farallon Izu-Nigata triple junction in the Late Jurassic and Early Cretaceous (M21–M10). *Tectonophysics* **155**, 345–364 (1988).
- Bunge, H.-P. & Grand, S. P. Mesozoic plate-motion history below the northeast Pacific Ocean from seismic images of the subducted Farallon slab. *Nature* **405**, 337–340 (2000).
- Ren, Y., Stutzmann, E., van der Hilst, R. D. & Besse, J. Understanding seismic heterogeneities in the lower mantle beneath the Americas from seismic tomography and plate tectonic history. *J. Geophys. Res.* **112**, B01302, <http://dx.doi.org/10.1029/2005JB004154> (2007).
- Decker, J. et al. *Geology of Southwestern Alaska* (eds Plafker, G. & Berg, H. C.) Vol. G-1, 285–310 (Geological Society of America, 1994).
- Poulton, T. P. et al. in *The Geological Atlas of the Western Canada Sedimentary Basin* Ch. 18 (2013); at [http://www.agi.ca/publications/abstracts/DIG\\_2008\\_0252.html](http://www.agi.ca/publications/abstracts/DIG_2008_0252.html).
- Saha, A., Basu, A. R., Wakabayashi, J. & Wortman, G. L. Geochemical evidence for a subducted infant arc in Franciscan high-grade-metamorphic tectonic blocks. *Geol. Soc. Am. Bull.* **117**, 1318–1335 (2005).
- Dickinson, W. R. Accretionary Mesozoic-Cenozoic expansion of the Cordilleran continental margin in California and adjacent Oregon. *Geosphere* **4**, 329–353 (2008).
- Ernst, W. G. Accretion of the Franciscan Complex attending Jurassic-Cretaceous geotectonic development of northern and central California. *Bull. Geol. Soc. Am.* **123**, 1667–1678 (2011).
- Leier, A. L. & Gehrels, G. E. Continental-scale detrital zircon provenance signatures in Lower Cretaceous strata, western North America. *Geology* **39**, 399–402 (2011).
- Johnston, S. T. et al. Yellowstone in Yukon: the Late Cretaceous Carmacks Group. *Geology* **24**, 997–1000 (1996).
- van de Zedde, D. M. A. & Wortel, M. J. R. Shallow slab detachment as a transient source of heat at midlithospheric depths. *Tectonics* **20**, 868–882 (2001).
- McDowell, F. W., Roldan-Quintana, J. & Connally, J. N. Duration of Late Cretaceous–early Tertiary magmatism in east-central Sonora, Mexico. *Geol. Soc. Am. Bull.* **113**, 521–531 (2001).
- González-León, C. M. et al. Stratigraphy, geochronology, and geochemistry of the Laramide magmatic arc in north-central Sonora, Mexico. *Geosphere* **7**, 1392–1418 (2011).
- Liu, L. et al. The role of oceanic plateau subduction in the Laramide orogeny. *Nature Geosci.* **3**, 353–357 (2010).
- Livaccari, R. F., Burke, K. & Sengor, A. M. C. Was the Laramide Orogeny related to subduction of an oceanic plateau? *Nature* **289**, 276–278 (1981).
- Enkin, R. J., Mahoney, J. B., Baker, J., Riesterer, J. & Haskin, M. L. Deciphering shallow paleomagnetic inclinations: 2. Implications from Late Cretaceous strata overlapping the Insular/Intermontane Superterrane boundary in the southern Canadian Cordillera. *J. Geophys. Res. B* **108**, 2186, <http://dx.doi.org/10.1029/2002JB001983> (2003).

42. Massey, N. Metchosin Igneous Complex, Southern Vancouver Island—ophiolite stratigraphy developed in an emergent island setting. *Geology* **14**, 602–605 (1986).
43. Wells, R. E. Reconsidering the origin and emplacement of Siletzia. *Geol. Soc. Am. Abstr. Prog.* **39** (4), 19 (2007).
44. Schmandt, B. & Humphreys, E. Seismically imaged relict slab from the 55 Ma Siletzia accretion to the northwest United States. *Geology* **39**, 175–178 (2011).
45. Shephard, G. E. *et al.* Testing absolute plate reference frames and the implications for the generation of geodynamic mantle heterogeneity structure. *Earth Planet. Sci. Lett.* **317–318**, 204–217 (2012).
46. Liu, L., Spasojević, S. & Gurnis, M. Reconstructing Farallon plate subduction beneath North America back to the Late Cretaceous. *Science* **322**, 934–938 (2008).
47. Steinberger, B., Torsvik, T. H. & Becker, T. W. Subduction to the lower mantle—a comparison between geodynamic and tomographic models. *Solid Earth* **3**, 415–432 (2012).
48. Zhong, S. & Gurnis, M. Mantle convection with plates and mobile, faulted plate margins. *Science* **267**, 838–843 (1995).
49. van der Meer, D. G., Torsvik, T. H., Spakman, W., van Hinsbergen, D. J. J. & Amaru, M. L. Intra-Panthalassa Ocean subduction zones revealed by fossil arcs and mantle structure. *Nature Geosci.* **5**, 215–219 (2012).
50. Torsvik, T. H., Müller, R. D., Van der Voo, R., Steinberger, B. & Gaina, C. Global plate motion frames: toward a unified model. *Rev. Geophys.* **46**, RG3004, <http://dx.doi.org/10.1029/2007RG000227> (2008).
51. Müller, R. D., Royer, J.-Y. & Lawver, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* **21**, 275–278 (1993).
52. Steinberger, B. & Torsvik, T. H. Absolute plate motions and true polar wander in the absence of hotspot tracks. *Nature* **452**, 620–623 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Müller for making available the plate reconstructions of ref. 22 before publication (for Fig. 1), as well as the Shatsky plateau reconstructions of ref. 39 for Fig. 3. We thank G. W. Ernst for a constructive review. The P-wave tomography model used here is available in ASCII format as part of the Auxiliary Materials for ref. 13 at <http://onlinelibrary.wiley.com/doi/10.1029/2010GC003421/supplinfo> or may be obtained from K.S. This is British Columbia Geological Survey contribution #2012-2.

**Author Contributions** K.S. generated the tomographic model, and integrated it with quantitative plate tectonic reconstructions in GPlates. M.G.M. provided the geological background and made the terrane maps of Fig. 3. Both authors contributed equally to developing the tectonic arguments and to the writing.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.S. ([sigloch@geophysik.uni-muenchen.de](mailto:sigloch@geophysik.uni-muenchen.de)) or M.G.M. ([Mitch.Mihalynuk@gov.bc.ca](mailto:Mitch.Mihalynuk@gov.bc.ca)).

## METHODS

The MEZ/ANG slab walls have been among the most robust features in global-scale body-wave tomography, starting with the work of Grand<sup>9–13,28</sup>. The deep end of the CR slab was already visible in some of the earlier studies. Its continuous upward connection to present-day Cascadia subduction was pointed out by ref. 15, hence its identification as a Farallon slab. The model on which we base our discussion here<sup>13</sup> is an inversion of P-wave observations recorded by North American broadband stations, using a cutting-edge waveform inversion technique (multi-frequency tomography) on a global, adaptive grid. Method discussion and formal resolution tests are presented in ref. 13. The higher resolution compared to global tomography models is largely due to densely spaced stations from the USArray experiment in the western half of the US, and waveforms recorded 2005–2008, which were not included in any of the above global models. Our calibration points for sinking rate, especially A3–A5, lie within the mantle subvolume that considerably benefits from resolution improvements afforded by the new USArray data.

We postulate that subduction into all slab walls imaged tomographically beneath North America<sup>13</sup> originated before the arrival of North America's western margin. Hence we must demonstrate sufficiently old slab ages, and cessation or flipping of subduction when the continent overrode the slabs. Plate reconstructions predict the timing of margin arrival above a slab, but only if slab and plate reconstructions<sup>12,21,22,50–52</sup> can be linked to the same absolute reference frame. (Technically we accomplish this with the free community palaeo-geographic information system GPlates<sup>23,24</sup> and a compilation of digitally published reference frames<sup>45</sup>.) Hence override predictions are correct to the extent that slab walls sank vertically, meaning that their  $x$ - $y$  locations since subduction are unchanged in a linked hotspot reference frame. Uncertainty is best quantified from slab wall geometry itself: deviation from vertical sinking probably did not exceed a wall's half-width (200–300 km), else such steep geometries could not have built up over a long time.

Uncertainties about absolute locations of North America's palaeo-margin arise from imperfections in plate reconstructions. Owing to terrane accretions and removals, there is also uncertainty about the shape and westward extent of North America's margin, compared to its present-day outlines. We discuss individual uncertainties in the Supplementary Information, and propagate them into cumulative uncertainties for the times at which North America's palaeo-margin overrode selected points A1–A5 on the palaeo-trenches. With relative uncertainties of only 10–15% (Table 1 and Supplementary Tables 1 and 2), this yields the old (Jura–Cretaceous) slab ages required to support intra-oceanic subduction.

Spatiotemporal predictions of trench override are verified by terrane observations as follows. Override should coincide with observable collision events, because buoyant island arcs or plateaus are overridden. Uncertainties on terrane observations are difficult to quantify, but particularly characteristic events (A1–A5) can nonetheless be singled out and used successfully for validation (Table 1). Sinking-rate calculations provide an additional plausibility check. Consistent results of 9–12 mm yr<sup>−1</sup> ( $\pm 1$ –2 mm yr<sup>−1</sup>) across all three slab walls show that this type of feature seems to sink rather predictably and evenly. If this is the case, then conversely the override prediction times that lead to the rate estimates should be adequate.

Sinking at about 10 mm yr<sup>−1</sup> is considered a lower-mantle sinking rate, but we obtain it as an average over both upper and lower mantle. This is explicable if a wall were to have sunk in steady state: upon subduction of its youngest end (Fig. 2b), its lower part would already have entered the viscous lower mantle and would have been setting the speed limit from below, which would have acted on the entire pile. Such a wall, widened to 4–6 times the lithospheric thickness (400–600 km), and sinking at about 10 mm yr<sup>−1</sup>, generates the same material throughput as the typical 40–60 mm yr<sup>−1</sup> of unbuckled plate convergence in the uppermost mantle<sup>16,20</sup>—a plausibility check that confirms the continuity of upper and lower mantle fluxes.

Sinking rates are obtained by dividing the imaged depth of a wall's shallowest end (depth reached since the end of subduction) by predicted time since the trench override. Besides timing uncertainty, a spatial uncertainty about the slab's true depth enters, as discussed quantitatively in Supplementary Fig. 2. These uncertainties are comparatively small, because upward truncations of the slab

walls are sharply imaged. This is also a striking visual feature (see, for example, Fig. 3g or Supplementary Fig. 2): along their lengths, MEZ and ANG show abrupt upward truncations in red to orange, yellow and green depth levels. This provides direct observational evidence for westward subduction—after trench override, there was no slab left to subduct. In contrast, the less-complete upward truncation of eastward-dipping CR around A5 (cyan colour level, a more localized slab window) indicates 'only' a larger terrane accretion event and trench step-back when the margin transitioned from intra-oceanic to Andean-style. Upward truncations shallow to the west (A1 at red level, A2 at yellow level, A3/A4 at green level, and A5 at cyan level), reflecting more and more recent ages for termination of subduction.

In summary, the observational uncertainties that we discuss in the Supplementary Information are:

(1) Uncertainties in plate reconstructions at the surface. When exactly did the North American palaeo-margin overlie a given point in the reference frame, for example, A1–A5? This includes relative reconstruction uncertainties (essentially ambiguities about the Atlantic Ocean opening; they are small, and neglected in our calculations); uncertainties about absolute reference frame (these are considerable; we attempt to quantify them by comparing different reference frames<sup>12,21,22,45,50–52</sup> (Supplementary Fig. 3); and uncertainties about the true westward extent of the palaeo-margin over time. The latter uncertainty is most difficult to quantify, since it requires knowledge about accreted terrane locations, which shifted over time. Our best guess of the uncertain area's extent is hatched in Fig. 1a and b and Supplementary Fig. 3a and b. Margin uncertainty is the biggest contributor to reconstruction uncertainties (typically 6–7 Myr), except for the oldest point A1, where absolute reference frame uncertainty dominates (Supplementary Table 2).

(2) Uncertainties in palaeo-trench locations, relative to the slab. Were trench points A1–A5 centred on the imaged slab walls, systematically offset to one side, or oscillating? We assume that the trenches ran centred, for lack of evidence to the contrary. A constant offset would hardly change the calculations, producing a strong correlation with margin uncertainty rather than independent uncertainties. Periodic trench advance and retreat—as buckling folds are being laid down—cannot be excluded, but we are unaware of observational evidence. Unless these oscillations significantly exceeded the half-width of the slab wall of 200–300 km (unlikely, given the tall slab piles), they would not dominate over margin uncertainty.

(3) Uncertainties about present-day slab depth (pertinent only to sinking-rate estimates). Evaluation of tomographic blur yields significantly smaller relative uncertainties than reconstruction errors (Supplementary Table 1). The two types of relative errors enter symmetrically into cumulative uncertainty on sinking rate.

Regarding our hypothesis of intra-oceanic trenches, the most important, qualitative assessment of uncertainty is that all the plate reconstructions considered<sup>21,22,45,50–52</sup> (with one exception<sup>12</sup> discussed below) agree that two ocean basins should have existed between North America and the MEZ/ANG slabs before 140 Myr ago (Fig. 1a and Supplementary Fig. 3a). These oceans were considerably wider than the uncertainty in margin extent, implying intra-oceanic subduction origins for all imaged slab walls, provided they are older than 140 Myr. That this is the case is shown by good agreement between predicted and geologically observed collision events (within the moderate error bars of Table 1). An extrapolation of sinking rates of about 10 mm yr<sup>−1</sup> to the lower ends of the slab walls at over 1,800 km depth implies that subduction originated at least 180 Myr ago.

The one reference frame<sup>12</sup> that does not predict wide ANG/MEZ ocean basins in the late Jurassic/early Cretaceous was explicitly designed to minimize their extents, by imposing a constraint that keeps North America's western margin stationary above the MEZ/ANG slabs (that is, an Andean-style margin is enforced a priori; SUB on Supplementary Fig. 3b). This is accomplished by introducing an additional degree of freedom, an *ad hoc*, otherwise non-observable, westward shift of the lithospheric shell relative to the lower mantle. In solving for this longitudinal shift, the method considers a global slab inventory, but for the times discussed here, the influence of the MEZ/ANG slabs dominates. Hence, this reference frame does not lend itself to evaluating the existence of the ANG/MEZ oceans.

# The genomes of four tapeworm species reveal adaptations to parasitism

Isheng J. Tsai<sup>1,2\*</sup>, Magdalena Zarowiecki<sup>1\*</sup>, Nancy Holroyd<sup>1\*</sup>, Alejandro Garcíarrubio<sup>3\*</sup>, Alejandro Sanchez-Flores<sup>1,3</sup>, Karen L. Brooks<sup>1</sup>, Alan Tracey<sup>1</sup>, Raúl J. Bobes<sup>4</sup>, Gladis Fragoso<sup>4</sup>, Edda Sciuotto<sup>4</sup>, Martin Aslett<sup>1</sup>, Helen Beasley<sup>1</sup>, Hayley M. Bennett<sup>1</sup>, Jianping Cai<sup>5</sup>, Federico Camicia<sup>6</sup>, Richard Clark<sup>1</sup>, Marcela Cucher<sup>6</sup>, Nishadi De Silva<sup>1</sup>, Tim A. Day<sup>7</sup>, Peter Deplazes<sup>8</sup>, Karel Estrada<sup>3</sup>, Cecilia Fernández<sup>9</sup>, Peter W. H. Holland<sup>10</sup>, Junling Hou<sup>5</sup>, Songnian Hu<sup>11</sup>, Thomas Huckvale<sup>1</sup>, Stacy S. Hung<sup>12</sup>, Laura Kamenetzky<sup>6</sup>, Jacqueline A. Keane<sup>1</sup>, Ferenc Kiss<sup>13</sup>, Uriel Koziol<sup>13</sup>, Olivia Lambert<sup>1</sup>, Kan Liu<sup>11</sup>, Xuenong Luo<sup>5</sup>, Yingfeng Luo<sup>11</sup>, Natalia Macchiaroli<sup>6</sup>, Sarah Nichol<sup>1</sup>, Jordi Paps<sup>10</sup>, John Parkinson<sup>12</sup>, Natasha Pouchkina-Stantcheva<sup>14</sup>, Nick Riddiford<sup>14,15</sup>, Mara Rosenzvit<sup>6</sup>, Gustavo Salinas<sup>9</sup>, James D. Wasmuth<sup>16</sup>, Mostafa Zamanian<sup>17</sup>, Yadong Zheng<sup>5</sup>, The *Taenia solium* Genome Consortium†, Xuepeng Cai<sup>5</sup>, Xavier Soberón<sup>3,18</sup>, Peter D. Olson<sup>14</sup>, Juan P. Laclette<sup>4</sup>, Klaus Brehm<sup>13</sup> & Matthew Berriman<sup>1</sup>

**Tapeworms (Cestoda) cause neglected diseases that can be fatal and are difficult to treat, owing to inefficient drugs.** Here we present an analysis of tapeworm genome sequences using the human-infective species *Echinococcus multilocularis*, *E. granulosus*, *Taenia solium* and the laboratory model *Hymenolepis microstoma* as examples. The 115- to 141-megabase genomes offer insights into the evolution of parasitism. Synteny is maintained with distantly related blood flukes but we find extreme losses of genes and pathways that are ubiquitous in other animals, including 34 homeobox families and several determinants of stem cell fate. Tapeworms have specialized detoxification pathways, metabolism that is finely tuned to rely on nutrients scavenged from their hosts, and species-specific expansions of non-canonical heat shock proteins and families of known antigens. We identify new potential drug targets, including some on which existing pharmaceuticals may act. The genomes provide a rich resource to underpin the development of urgently needed treatments and control.

Echinococcosis (hydatid disease) and cysticercosis, caused by the proliferation of larval tapeworms in vital organs<sup>1</sup>, are among the most severe parasitic diseases in humans and account for 2 of the 17 neglected tropical diseases prioritized by the World Health Organization<sup>2</sup>. Larval tapeworms can persist asymptotically in a human host for decades<sup>3</sup>, eventually causing a spectrum of debilitating pathologies and death<sup>1</sup>. When diagnosed, the disease is often at an advanced stage at which surgery is no longer an option<sup>4</sup>. Tapeworm infections are highly prevalent worldwide<sup>5</sup>, and their human disease burden has been estimated at 1 million disability-adjusted life years, comparable with African trypanosomiasis, river blindness and dengue fever. Furthermore, cystic echinococcosis in livestock causes an annual loss of US\$2 billion<sup>6</sup>.

Tapeworms (Platyhelminthes, Cestoda) are passively transmitted between hosts and parasitize virtually every vertebrate species<sup>7</sup>. Their morphological adaptations to parasitism include the absence of a gut, a head and light-sensing organs, and they possess a unique surface ( tegument) that is able to withstand host-stomach acid and bile but is still penetrable enough to absorb nutrients<sup>7</sup>.

Tapeworms are the only one of three major groups of worms that parasitize humans, the others being flukes (Trematoda) and round worms (Nematoda), for which no genome sequence has been available so far. Here we present a high-quality reference tapeworm genome of a human-infective fox tapeworm *Echinococcus multilocularis*. We also present the genomes of three other species, for comparison; *E. granulosus* (dog tapeworm), *Taenia solium* (pork tapeworm), both of which infect humans, and *Hymenolepis microstoma* (a rodent tapeworm and laboratory model for the human parasite *Hymenolepis nana*). We have mined the genomes to provide a starting point for developing urgently needed therapeutic measures against tapeworms and other parasitic flatworms. Access to the complete genomes of several tapeworms will accelerate the pace at which new tools and treatments to combat tapeworm infections can be discovered.

## The genomes and genes of tapeworms

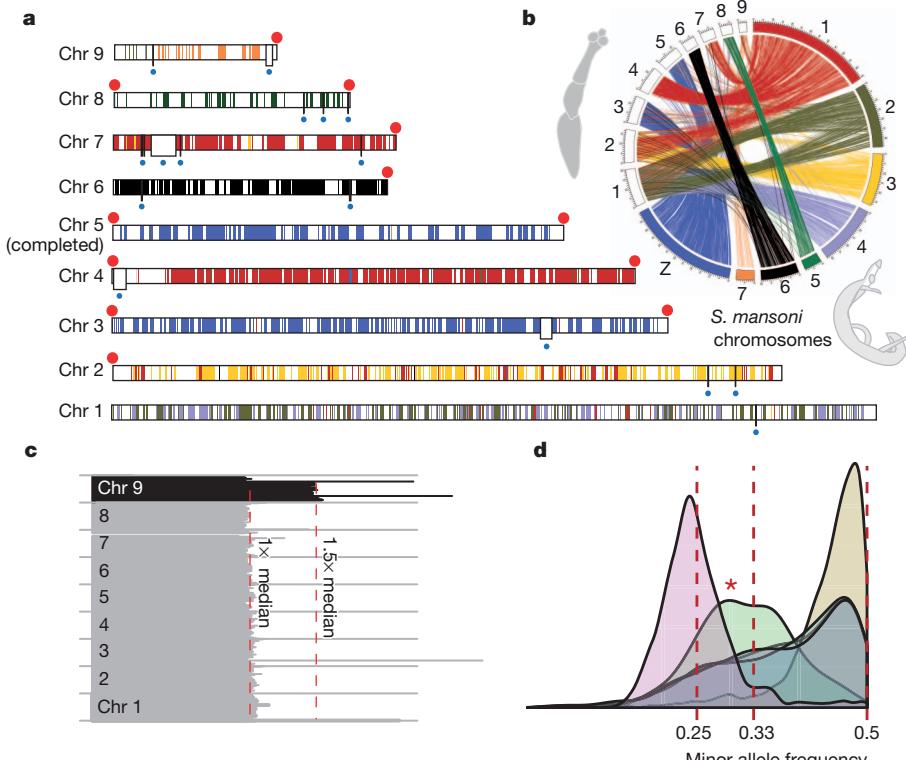
The *E. multilocularis* genome assembly was finished manually (Supplementary Information, section 2), producing a high-quality reference

<sup>1</sup>Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>2</sup>Division of Parasitology, Department of Infectious Disease, Faculty of Medicine, University of Miyazaki, Miyazaki 889-1692, Japan. <sup>3</sup>Institute of Biotechnology, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. <sup>4</sup>Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México. <sup>5</sup>State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No. 1 Xujiaoping, Chengguan District, Lanzhou 730046, Gansu Province, China. <sup>6</sup>Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG Buenos Aires, Argentina.

<sup>7</sup>Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA. <sup>8</sup>Institute of Parasitology, Vetsuisse Faculty, University of Zürich, Winterthurerstrasse 266a, CH-8057 Zürich, Switzerland. <sup>9</sup>Cátedra de Inmunología, Facultad de Química, Universidad de la República. Avenida Alfredo Navarro 3051, piso 2, Montevideo, CP 11600, Uruguay. <sup>10</sup>Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. <sup>11</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, No. 7 Beitucheng West Road, Chaoyang District, Beijing 100029, China. <sup>12</sup>Department of Biochemistry & Molecular and Medical Genetics, University of Toronto, Program in Molecular Structure and Function, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. <sup>13</sup>University of Würzburg, Institute of Hygiene and Microbiology, D-97080 Würzburg, Germany. <sup>14</sup>Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. <sup>15</sup>Department of Zoology, School of Natural Sciences and Regenerative Medicine Institute (REMDI), National University of Ireland Galway, University Road, Galway, Ireland. <sup>16</sup>Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary T2N 4Z6, Canada. <sup>17</sup>Institute of Parasitology, McGill University, 2111 Lakeshore Road, Ste Anne de Bellevue, Quebec H9X 3V9, Canada.

<sup>\*</sup>These authors contributed equally to this work.

<sup>†</sup>Lists of participants and their affiliations appear at the end of the paper.



**Figure 1 | Genome of *E. multilocularis*.** **a**, The nine assembled chromosomes (Chr 1–Chr 9) of *E. multilocularis* with telomeres (red dots). Physical gaps in the sequence assembly (white boxes with blue dot beneath) are bridged by optical map data. Each colour segment is defined as an array of at least three genes that each has a single orthologous counterpart on one *S. mansoni* chromosome, regardless of their locations on the chromosome. **b**, One-to-one orthologues connecting *E. multilocularis* and *S. mansoni* chromosomes. **c**, Distribution of normalized genome coverage on isolate GT10/2. Each horizontal line depicts median coverage of 100-kb windows normalized against the mean coverage for the genome ( $130 \times$ ). Even coverage was observed across the first eight chromosomes, but  $1.5 \times$  coverage of chromosome 9 indicates trisomy. Similar plots for other isolates are shown in Supplementary Fig. 3.1. **d**, Distribution of minor allele frequency (MAF) of heterozygous sites in five isolates of *E. multilocularis* (plot for individual isolates in Supplementary Fig. 3.1), identified by mapping sequencing reads against the assembled chromosome consensus sequences. At each site, the proportion of bases that disagree with the reference is counted. For four isolates, the MAF peaks at around 0.5, indicative of diploidy, whereas JAVA05/1 peaks at 0.25, suggesting tetraploidy. Chromosome 9 of GT10/2 is plotted separately (marked by asterisk) from chromosomes 1 to 8, and the MAF display a clear departure of 0.5 and peaks around 0.33, consistent with a trisomy.

genome in which 89% of the sequence is contained in 9 chromosome scaffolds that have only 23 gaps (Supplementary Table 1.2). One chromosome is complete from telomere to telomere, and 13 of the expected 18 telomeres are joined to scaffolds (Fig. 1a). This quality and completeness is comparable to that of the first publications of *Caenorhabditis elegans* and *Drosophila melanogaster* genomes<sup>8,9</sup>. The 115- to 141-megabase (Mb) nuclear tapeworm genomes were sequenced using several high-throughput technologies (Supplementary Table 1.1). The tapeworm genomes are approximately one-third of the size of the genome of their distant flatworm relative, the blood fluke *Schistosoma mansoni*<sup>10</sup>, mainly because it has fewer repeats (Supplementary Information, section 3). By sequencing several isolates of *E. multilocularis* (Supplementary Table 3.2), we revealed tetraploidy in protoscoleces of one isolate, and a trisomy of chromosome 9 (the smallest chromosome, and possibly the only one for which a trisomy is tolerated) transiently exhibited in protoscoleces and metacercostomes from two different isolates (Fig. 1c, d and Supplementary Figs 3.1, 3.2 and 3.3), consistent with previous observations of karyotype plasticity in flatworms<sup>11</sup>.

Aided by deep transcriptome sequencing from multiple life-cycle stages, we identified 10,231 to 12,490 putative genes per genome (Supplementary Table 5.5). Similar to the genome of *S. mansoni*<sup>12</sup>, distinct ‘micro-exon genes’ are present in tapeworm genomes, with multiple internal exons that are small (typically less than 36 bases) and divisible by 3 (Supplementary Information, section 5). To identify gene gain and loss in tapeworms, orthologous relationships were predicted between tapeworms and eight other species (Fig. 2). Although gene order has been lost, ancient chromosomal synteny is preserved among parasitic flatworms (Fig. 1b and Supplementary Table 7.3). Two chromosomes in *E. multilocularis* (Fig. 1a, b) correspond to the *S. mansoni* Z sex chromosome. Schistosomes are unusual among flatworms in that they have sexual dimorphism, but how common ancestors of both tapeworms and flukes evolved into female-heterogametic parasites, like *S. mansoni*, remains to be elucidated.

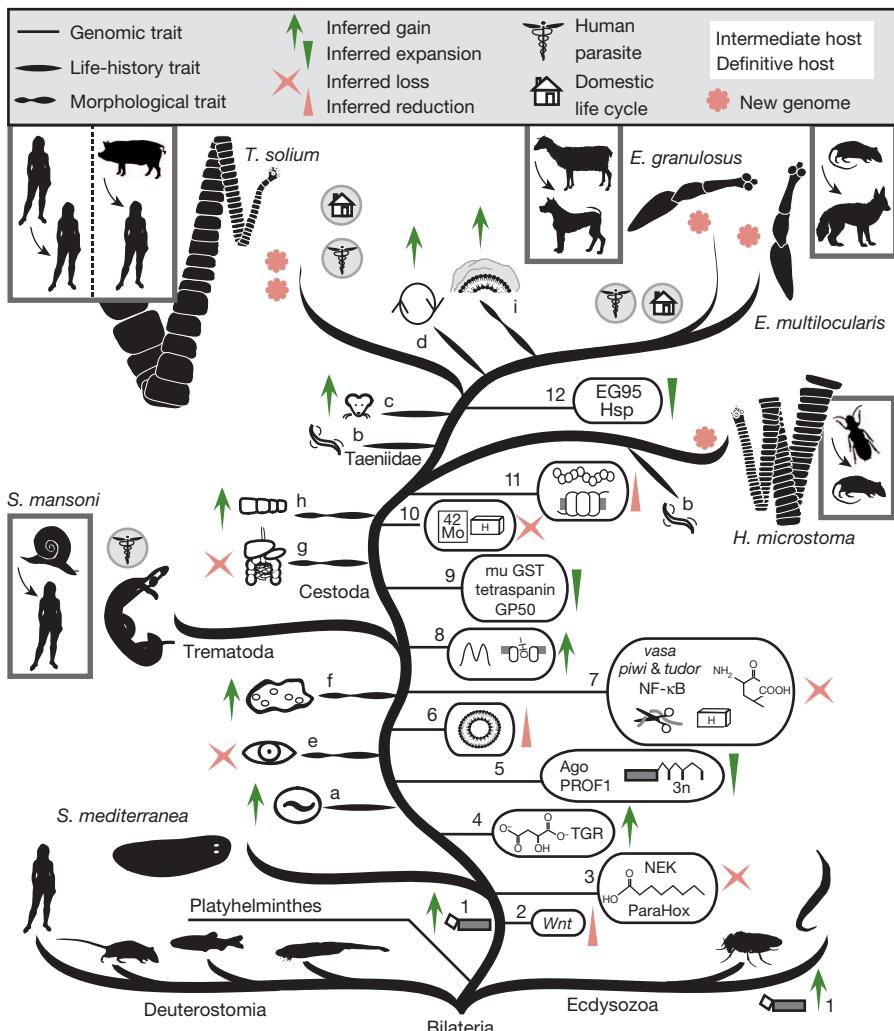
Genome-wide identification of polycistrons in tapeworms shows that there are 308 putative polycistrons in *E. multilocularis*, with the largest containing 4 genes. The internal gene order within *E. multilocularis* polycistrons is largely the same as in *T. solium* and *H. microstoma* (Supplementary Table 6.5), and—to some extent—as in flukes; 39% of *S. mansoni* orthologues of genes within *E. multilocularis* polycistrons retain colinearity. Of these *S. mansoni* genes, 40% have transcriptome evidence supporting their polycistronic transcription<sup>10</sup>, demonstrating further that gene order in polycistrons is highly conserved over long evolutionary time<sup>13</sup> ( $P < 0.0001$ , Supplementary Information, section 6).

Polycistrons are resolved into individual coding transcripts using spliced-leader *trans-splicing*, but spliced-leader *trans-splicing* also occurs in genes outside of polycistrons. Using deep transcriptome sequencing (RNA-seq) we found evidence of spliced-leader *trans-splicing* in approximately 13% of *E. multilocularis* genes (Supplementary Table 6.2), less than the 70% observed in *C. elegans*<sup>14</sup> and 58% in a tunicate<sup>15</sup>.

### Specialized metabolism and detoxification

The high-confidence gene sets reveal extensive reductions in overall metabolic capability and an increased ability to absorb nutrients, compared to that of other animals (Figs 2 and 3, and Supplementary Information, section 9). Their main energy source, carbohydrates, can be catabolized by aerobic respiration or by two complementary anaerobic pathways; the lactate fermentation and malate dismutation pathways. The parasiticidal effects of mitochondrial fumarate reductase inhibitors have been demonstrated *in vitro*, suggesting that the malate dismutation pathway would be an effective target for the development of novel therapeutics<sup>16</sup>.

Tapeworms, like flukes, lack the ability to synthesize fatty acids and cholesterol *de novo*<sup>17,18</sup>. Instead, they scavenge essential fats from the host using fatty acid transporters and lipid elongation enzymes (Supplementary Table 9.2), as well as several tapeworm-specific gene families, such as fatty acid binding protein (FABP) and the apolipoprotein



**Figure 2 | Evolution of tapeworm parasitism.** Phylogeny of the main branches of Bilateria; Ecdysozoa (including fruitflies and nematodes), Deuterostomia (including lancelet, zebrafish, mice and humans), and Lophotrochozoans (including Platyhelminthes (flatworms)) (based on phylogeny in Supplementary Fig. 7.1). The gains and losses of life-cycle traits for these parasitic flatworms include the evolution of endoparasitism (a), passive transmission between hosts (b), acquisition of vertebrate intermediate host (c), ability to proliferate asexually in intermediate host (d). Morphological traits that have evolved include the loss of eye cups (e), gain of neodermatan syncytial epithelia (f), loss of gut (g), segmentation of body plan (h), and changes in the laminated layer (to contain specialized apomucins; i). Gains and losses of genomic traits include spliced-leader trans-splicing (1), loss of *Wnt* genes (2), loss of NEK kinases, fatty acid biosynthesis and ParaHox genes (3), anaerobic metabolic ability through the malate dismutation/rohydroquinone pathway, merger of glutaredoxin and thioredoxin reductase to thioredoxin glutathione reductase (TGR) (4), evolution of tapeworm- and fluke-specific Argonaute (Ago) family, micro exon genes (MEGs) and PROF1 GPCRs (5), loss of peroxisomal genes (6), and complete loss of *vasa*, *tudor* and *piwi* genes, NF- $\kappa$ B pathway, loss of 24 homeobox gene families (indicated by 'H'), metabolic proteases and amino acid biosynthesis (7). In tapeworms, gains and losses of genomic traits include innovation of bimodal intron distribution and novel fatty acid transporters (8), expansion of mu-class glutathione S-transferases, GP50 antigens and tetraspanins (9), loss of the molybdopterin biosynthesis pathway, loss of 10 homeobox gene families (10), fewer GPCRs and fewer neuropeptides encoded by each protopeptide (11), and expansion of heat shock proteins (Hsp) and species-specific antigens (12).

antigen B (Supplementary Information, section 8). Uptake of fatty acids seems to be crucial in *Echinococcus* spp. metacestodes, in which both FABP and antigen B gene families are among the most highly expressed genes<sup>19</sup> (Supplementary Table 5.7). Tapeworms and flukes have lost many genes associated with the peroxisome (Supplementary Information, section 8), an organelle in which fatty acid oxidation occurs, and may lack peroxisomes altogether, as seen in several other parasites<sup>20</sup>.

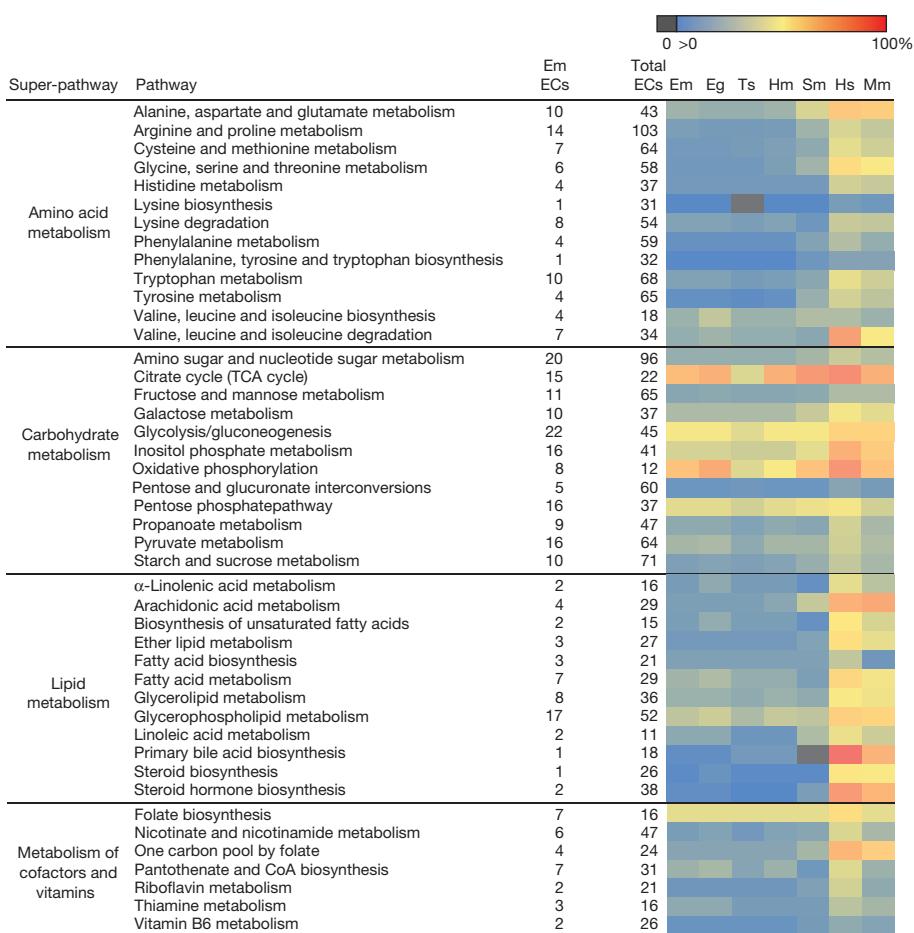
Compared with other animals, *S. mansoni* has a reduced ability to synthesize amino acids<sup>17</sup>. In tapeworms, this capacity is reduced further, with serine and proline biosynthesis enzymes absent from *E. multilocularis* (Fig. 3 and Supplementary Information, section 9). Many enzymes in the molybdopterin biosynthesis pathway seemed to be lost in tapeworms, along with enzymes that use molybdopterin as a cofactor. The ability to utilize molybdenum in enzymatic reactions was believed to be present in all animals<sup>21</sup>, but has been lost in some eukaryotic parasites<sup>22</sup>.

Differences in the detoxification systems between tapeworms and their mammalian hosts may be exploited for drug design (Supplementary Information, section 9). We found that, like flukes<sup>23</sup>, tapeworms typically have only one cytochrome P450 gene, suggesting that their ability to oxidize many xenobiotics and steroids is substantially lower than that of their hosts. Uniquely, tapeworms and flukes have merged two key enzymatic functions for redox homeostasis in one single enzyme: thioredoxin glutathione reductase (TGR). TGR is an essential gene and validated drug target in flukes<sup>24</sup>. Downstream of TGR we find an unexpected diversity of thioredoxins, glutaredoxins

and mu-class glutathione S-transferases (GSTs) (Supplementary Table 9.3). The GST expansion suggests that tapeworms would be able to water-solubilize and excrete a large range of hydrophobic compounds, which may add complexity to the pharmacokinetics of drugs.

### Homeobox gene loss

Homeobox genes are high-level transcription factors that are implicated in the patterning of body plans in animals. Across parasitic flatworms, the homeobox gene numbers are extensively reduced (Supplementary Table 10.1). Most bilaterian invertebrates have a conserved set of approximately 100 homeobox genes (for example, 92 conserved in *C. elegans*, 102 in *D. melanogaster*, and 133 in the lancelet)<sup>25</sup>. Of the 96 homeobox gene families that are thought to have existed at the origin of the Bilateria, 24 are not present in tapeworms and flukes, and a further 10 were lost in tapeworms, making their complement by far the most reduced of any studied bilaterian animal<sup>25</sup>. Among the tapeworm-specific gene losses are gene families involved in neural development (*mnx*, *pax3/7*, *gbx*, *hbn* and *rax*). This is somewhat surprising considering that tapeworms possess a well-developed nervous system, albeit with reduced sensory input and cephalization. Tapeworms also lack the ParaHox genes (*gsx*, *pdx*, *cdx*) ancestrally involved in specification of a through-gut<sup>26,27</sup>, although these seem to have been lost before the tapeworm gut was lost. Other conserved genes found in bilaterian developmental pathways such as Hedgehog and Notch were found to be present and intact, although the *Wnt* complement is greatly reduced compared to the ancestral (spiralian) complement of 12 *Wnt* ligands<sup>28</sup> (Supplementary Table 10.2).



**Figure 3 | Conservation of individual metabolic pathways.** Heatmap showing the conservation of individual metabolic pathways for *E. multilocularis* (Em), *E. granulosus* (Eg), *T. solium* (Ts), *H. microstoma* (Hm) and *S. mansoni* (Sm) compared to those of humans (Hs) and mice (Mm). Each row indicates an individual metabolic pathway grouped by their superclass membership (defined by KEGG (Kyoto Encyclopedia of Genes and Genomes)). Coloured tiles indicate the level of conservation (percentage of enzymes detected) of each pathway within each species. KEGG pathways with insufficient evidence (that is, containing only one enzyme) in *E. multilocularis* have been removed. CoA, coenzyme A; EC, enzyme commission number; TCA, tricarboxylic acid cycle.

## Stem cell specializations

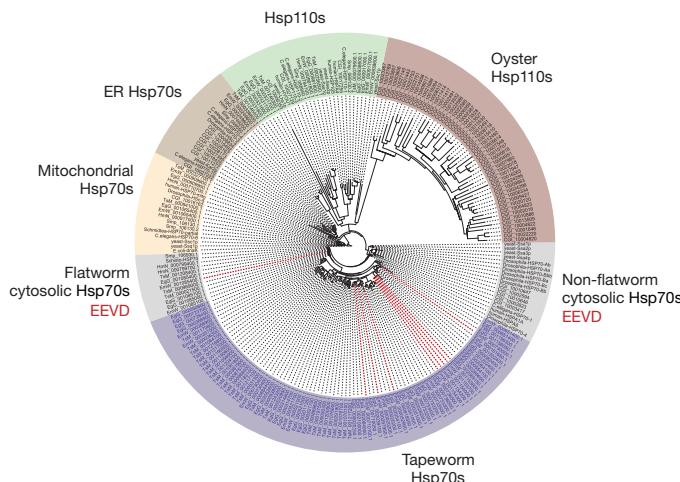
Extreme regenerative capability and developmental plasticity, mediated by ever-present somatic stem cells (neoblasts), have made flatworms popular models for stem cell research<sup>29</sup>. All multicellular organisms rely on stem cells for proliferation and growth, so it is remarkable that tapeworms and flukes appear to lack the ubiquitous stem cell marker gene *vasa* (Supplementary Information, section 11). Instead tapeworms have two copies of another dead-box helicase (PL10), which we propose may have taken over some of the functions of *vasa* (Supplementary Fig. 11.1). Tapeworms and flukes are also missing the *piwi* gene subfamily and *piwi*-interacting *tudor*-domain containing proteins. The *piwi* genes belong to a subfamily of genes encoding argonaute proteins, and we also found that tapeworms have a new subfamily of argonaute proteins (Supplementary Fig. 11.2) that may bind a newly discovered potential small RNA precursor<sup>30</sup>. Both *piwi* and *vasa* are usually essential in regulating the fate of germline stem cells in animals, and *vasa* suppression usually leads to infertility or death<sup>31</sup>. These findings suggest that stem-cell-associated pathways in parasitic flatworms may be highly modified.

## Specialization of the tapeworm proteome

We sought to identify novel and expanded gene families in tapeworms, and found many frequently occurring novel domains involved in cell–cell adhesion and the formation of the tegument (Supplementary Information, section 8). For example, several novel domains are found on the ectodomain of cadherins (Supplementary Information, section 8), and tapeworms have proportionally more tetraspanin copies (30–36) (Supplementary Table 12.1) than the highly expanded repertoires of fruitflies and zebrafish<sup>32</sup>. The acellular carbohydrate-rich laminated layer, which coats the outside of *Echinococcus* metacestodes, is a unique genus-specific trait and one of the

few morphological traits that differ between the very closely related species *E. granulosus* and *E. multilocularis*. We identified corresponding species differences in an *Echinococcus*-specific apomucin family (Supplementary Fig. 12.1), an important building block of the laminated layer<sup>33</sup>. One particular copy is highly differentiated between the two species (non-synonymous to synonymous substitution ratio of  $>1$ ) and is the fifth most highly expressed in the metacestode stage of *E. multilocularis* (Supplementary Table 5.7). Galactosyltransferases that probably decorate the apomucins with galactose residues, the predominant sugar of laminated layer glycans, are similarly diverged<sup>33</sup> (Supplementary Information, section 8). Approximately 20% of the genes are exclusive to tapeworms, and these include many highly expressed antigen families, such as antigen B, the glycosylphosphatidyl-inositol (GPI)-anchored protein GP50 (ref. 34), and the vaccine target EG95 (ref. 35) (Supplementary Table 12.4).

One of the most striking gene family expansions in the tapeworm genomes is the heat shock protein 70 (Hsp70) family. Phylogenetic analysis revealed independent and parallel expansions in both the Hsp110 and the cytosolic Hsp70 clades (Fig. 4). Several examples of expansions exist at various clades of Hsp70 in other systems, including Hsp110 expansions in oysters (to cope with temperature) and in cancer cells (to cope with proteotoxic stress)<sup>36,37</sup>. *Echinococcus* and *T. solium* have the highest number of gene expansions in the cytosolic Hsp70 clade. These expansions seem to have occurred independently in each species, and have resulted in 22 to 32 full copies in each species (*Echinococcus* and *T. solium*) compared to 6 copies in fruitflies and 2 in humans (Fig. 4). This expanded clade lacks classical cytosolic Hsp70 features (a conserved EEVD motif for substrate binding and a GGMP repeat unit), and whereas the canonical cytosolic *hsp70* genes are constitutively expressed in different life-cycle stages, the non-canonical genes show almost no expression, suggesting a putative



**Figure 4 | Heat shock protein 70 expansions in tapeworms.** Rooted tree of Hsp70 sequences from tapeworms and the eight comparator species used in this study, with additional sequences from baker's yeast *Saccharomyces cerevisiae*, and the Pacific oyster *Crassostrea gigas* (a non-flatworm example of a lophotrochozoan with a recently reported Hsp70 expansion). Different Hsp70 subfamilies are shown in different colours. Dotted red lines, *E. multilocularis* hsp70 genes that are located in the subtelomeres. EEVD, the conserved carboxy-terminal residues of a canonical cytosolic Hsp70; ER Hsp70, endoplasmic reticulum Hsp70.

contingency role in which individual copies of the expanded family are only highly expressed under certain conditions (Supplementary Fig. 12.2). At least 40% of *E. multilocularis* hsp70-like genes are found within the subtelomeric regions of chromosomes, including the extreme case of chromosome 8 in which eight copies (including pseudogenes) are located in the subtelomere (Supplementary Table 12.2). No other genes are over-represented in these regions. Although Hsp70 proteins have been found in excretory–secretory products of tapeworms<sup>38</sup>, it remains to be determined whether the non-canonical Hsps have a host-interacting role or whether telomere proximity is important for their function or expression.

## Novel drug targets

Tapeworm cysts are treated by chemotherapy or surgical intervention depending on tapeworm species, patient health and the site of the cyst. The only widely used drugs to treat tapeworm cysts are benzimidazoles<sup>39</sup> that, owing to considerable side effects, are administered at parasitostatic rather than parasiticidal concentrations<sup>40</sup>. Novel targets and compound classes are therefore urgently needed.

To identify new potential drug targets, we surveyed common targets of existing pharmaceuticals; kinases, proteases, G-protein-coupled receptors (GPCRs) and ion channels<sup>41</sup>. We identified approximately 250 to 300 new protein kinases (Supplementary Table 13.1), and these cover most major classes (Supplementary Information, section 13). We also identified 151 proteases and 63 peptidase-like proteins in *E. multilocularis*, a repertoire of similar diversity to *S. mansoni*, and found that, like *S. mansoni*, *E. multilocularis* has strongly reduced copy numbers compared to those of other animals (Supplementary Table 13.9). Many successful anthelmintic drugs target one of several different forms of neural communication<sup>41</sup>. We therefore mapped the signalling pathways of the serotonin and acetylcholine neurotransmitters, predicted conserved and novel neuropeptides (Supplementary Table 13.6), and classified more than 60 putative GPCRs (Supplementary Table 13.2) and 31 ligand-gated ion channels (Supplementary Table 13.4). A voltage-gated calcium channel subunit<sup>42</sup>—the proposed target of praziquantel—is not expressed in cysts and thus provides a putative explanation for the drug's low efficacy.

We searched databases for potential features for target selection, including compounds associated with protein targets and expression in the clinically relevant metacestode life-stage, and using this information we assigned weights to rank the entire proteomes (Supplementary Table 13.10). We identified 1,082 *E. multilocularis* proteins as potential targets, and of these, 150 to 200 with the highest scores have available chemical leads (known drug or approved compounds).

Acetylcholinesterases, which are inhibited by mefloquine (an antimalarial that reduces egg production in *S. mansoni*), are high on the list of potential targets<sup>43</sup>. However, acetylcholinesterase transcription in tapeworm cysts is low, possibly limiting their suitability. After filtering to remove targets with common substrates rather than inhibitors, the top of the list includes several homologues of targets for

**Table 1 | Top 20 promising targets in *E. multilocularis***

| Target category       | Target  | Action                        | Expression | Drug                                   | Rank |
|-----------------------|---|-------------------------------|------------|--|------|
| Current targets       | Tubulin β-chain                                   | Cytoskeleton                  | M,A        | Albendazole                            | 406  |
| Potential target      | Voltage-dependent calcium channel                 | Ion transport                 | -          | Praziquantel                           | 277  |
| Top predicted targets | Thioredoxin glutathione reductase (TGR)           | Detoxification                | M,A        | Experimental compounds                 | 277  |
|                       | Fatty acid amide hydrolase                        | Bioactive lipid catabolism    | M          | Thiopental, propofol                   | 1    |
|                       | Adenine nucleotide translocator                   | Mitochondrial ATP export      | M          | Clodronate                             | 2    |
|                       | Inosine 5' monophosphate dehydrogenase            | Purine biosynthesis           | M          | Mycophenolic acid, ribavirin           | 3    |
|                       | Succinate semialdehyde dehydrogenase              | GABA catabolism               | M          | Chlormerodrin                          | 3    |
|                       | Ribonucleoside diphosphate reductase              | Purine biosynthesis           | M,A        | Motexafin gadolinium                   | 5    |
|                       | Casein kinase II                                  | Cell-cycle regulating kinase  | M,A        | Experimental compounds                 | 6    |
|                       | Hypoxanthine guanine phosphoribosyltransferase    | Purine biosynthesis           | M,A        | Azathioprine                           | 8    |
|                       | Glycogen synthase kinase 3                        | Multiple signalling pathways  | M,A        | Lithium                                | 8    |
|                       | Proteasome subunit                                | Protein degradation           | M,A        | Bortezomib                             | 16   |
|                       | Calmodulin  | Transduces calcium signals    | M,A        | Trifluoperazine                        | 19   |
|                       | FK506 binding protein                             | Protein folding               | M,A        | Pimecrolimus                           | 19   |
|                       | UMP-CMP kinase                                    | Phosphorylases                | M          | Gemcitabine                            | 39   |
|                       | Na <sup>+</sup> /K <sup>+</sup> ATPase            | Ion transport                 | M          | Artemether                             | 42   |
|                       | Carbonic anhydrase II                             | Acidity control               | M          | Multiple (for example, Methazolamide)  | 42   |
|                       | NADH dehydrogenase subunit 1                      | Energy metabolism             | M          | Multiple (for example, Methoxyflurane) | 42   |
|                       | Translocator protein                              | Multiple functions            | M,A        | Multiple (for example, Lorazepam)      | 42   |
|                       | Elongation factor 2                               | Translation                   | M,A        | Experimental compounds                 | 54   |
|                       | Cathepsin B                                       | Protease                      | M          | Experimental compounds                 | 55   |
|                       | Dual-specificity mitogen activated protein kinase | Signalling, activation of p38 | M          | Experimental compounds                 | 56   |
|                       | Purine nucleoside phosphorylase                   | Purine metabolism             | M,A        | Didanosine                             | 63   |

A, adult; M, metacestode. Rank is sorted starting from the highest overall score; proteins with tied scores have the same rank. For current targets, the rank is only reported from the highest-scoring protein family member. For full scores and information please see Supplementary Table 13.10.

cancer chemotherapy, including casein kinase II, ribonucleoside reductase, UMP-CMP kinase and proteasome subunits (Table 1). The challenges of inhibiting cancer tumours and metacestodes (particularly those of *E. multilocularis*) with drugs are in some ways similar; both show uncontrolled proliferation, invasion and metastasis, and are difficult to kill without causing damage to the surrounding tissue. Therefore, metacestodes may be vulnerable to similar strategies as cancer; suppression of mitosis, induction of apoptosis and prevention of DNA replication. In fact, the anthelmintic medicines niclosamide, mebendazole and albendazole have already been shown to inhibit cancer growth<sup>44</sup>.

## Conclusion

Tapeworms were among the first known parasites of humans, recorded by Hippocrates and Aristotle in ~300 BC (ref. 45), but a safe and efficient cure to larval tapeworm infection in humans has yet to be found. These genomes provide hundreds of potential drug targets that can be tested using high-throughput drug screenings that were made possible by recent advances in axenic and cell culturing techniques<sup>39,46,47</sup>. Flatworms display an unusually high degree of developmental plasticity. In this study, the high level of sequence completion enabled both gene losses and gains to be accurately determined, and has shown how this plasticity has been put to use in the evolution of tapeworms.

## METHODS SUMMARY

Genome sequencing was carried out using a combination of platforms. RNA sequencing was performed with Illumina RNA-seq protocols (for *E. multilocularis*, *E. granulosus* and *H. microstoma*) or capillary sequencing of full-length complementary DNA libraries (*T. solium*). The complete genome annotation is available at <http://www.genedb.org>. The tapeworm genome projects were registered under the INSDC project IDs PRJEB122 (*E. multilocularis*), PRJEB121 (*E. granulosus*), PRJEB124 (*H. microstoma*) and PRJNA16816 (*T. solium*). Sequence data for *T. solium* isolate (from Mexico) were used for all orthologue comparisons, but results relating to gene gains and losses were reconciled against an additional sequenced isolate from China (unpublished). All experiments involving jirds (laboratory host of *E. multilocularis*) were carried out in accordance with European and German regulations relating to the protection of animals. Ethical approval of the study was obtained from the ethics committee of the government of Lower Franconia (621-2531.01-2/05). Experiments with dogs (host of *E. multilocularis* sample RNA-seq ERS018054) were conducted according to the Swiss guidelines for animal experimentation and approved by the Cantonal Veterinary Office of Zurich prior to the start of the study, and were carried out with facility-born animals at the experimental units of the Vetsuisse Faculty in Zurich (permission numbers 40/2009 and 03/2010). A licensed hunter hunted the fox (host of *E. multilocularis* sample RNA-seq ERS018053) during the regular hunting season. *Hymenolepis* parasites were reared using laboratory mice in accordance with project license PPL 70/7150, granted to P.D.O. by the UK Home Office.

Received 9 November 2012; accepted 21 February 2013.

Published online 13 March 2013.

- Garcia, H. H., Moro, P. L. & Schantz, P. M. Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. *Curr. Opin. Infect. Dis.* **20**, 489–494 (2007).
- World Health Organization. *Neglected Tropical Diseases* ([http://www.who.int/neglected\\_diseases/diseases/en/](http://www.who.int/neglected_diseases/diseases/en/)) (2012).
- Eckert, J. & Deplazes, P. Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. *Clin. Microbiol. Rev.* **17**, 107–135 (2004).
- Brunetti, E., Kern, P. & Vuitton, D. A. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Trop.* **114**, 1–16 (2010).
- Budke, C. M., White, A. C., Jr & Garcia, H. H. Zoonotic larval cestode infections: neglected, neglected tropical diseases? *PLoS Negl. Trop. Dis.* **3**, e319 (2009).
- Torgerson, P. R. & Macpherson, C. N. The socioeconomic burden of parasitic zoonoses: global trends. *Vet. Parasitol.* **182**, 79–95 (2011).
- Burton, J., Bogitsh, C. E. C. & Oeltmann, T. N. *Human parasitology*. 4th edn (Academic Press, 2012).
- Adams, M. D. et al. The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Protasio, A. V. et al. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **6**, e1455 (2012).
- Špakulová, M., Orosova, M. & Mackiewicz, J. S. Cytogenetics and chromosomes of tapeworms (Platyhelminthes, Cestoda). *Adv. Parasitol.* **74**, 177–230 (2011).
- DeMarco, R. et al. Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. *Genome Res.* **20**, 1112–1121 (2010).
- Qian, W. & Zhang, J. Evolutionary dynamics of nematode operons: easy come, slow go. *Genome Res.* **18**, 412–421 (2008).
- Allen, M. A., Hillier, L. W., Waterston, R. H. & Blumenthal, T. A global analysis of *C. elegans* trans-splicing. *Genome Res.* **21**, 255–264 (2011).
- Matsumoto, J. et al. High-throughput sequence analysis of *Ciona intestinalis* SL trans-spliced mRNAs: alternative expression modes and gene function correlates. *Genome Res.* **20**, 636–645 (2010).
- Matsumoto, J. et al. Anaerobic NADH-fumarate reductase system is predominant in the respiratory chain of *Echinococcus multilocularis*, providing a novel target for the chemotherapy of alveolar echinococcosis. *Antimicrob. Agents Chemother.* **52**, 164–170 (2008).
- Berriman, M. et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358 (2009).
- Frayha, G. J. Comparative metabolism of acetate in the taeniid tapeworms *Echinococcus granulosus*, *E. multilocularis* and *Taenia hydatigena*. *Comp. Biochem. Physiol. B* **39**, 167–170 (1971).
- Obal, G. et al. Characterisation of the native lipid moiety of *Echinococcus granulosus* antigen B. *PLoS Negl. Trop. Dis.* **6**, e1642 (2012).
- Kaasch, A. J. & Joiner, K. A. Targeting and subcellular localization of *Toxoplasma gondii* catalase. Identification of peroxisomes in an apicomplexan parasite. *J. Biol. Chem.* **275**, 1112–1118 (2000).
- Schwarz, G. & Mendel, R. R. Molybdenum cofactor biosynthesis and molybdenum enzymes. *Annu. Rev. Plant Biol.* **57**, 623–647 (2006).
- Zhang, Y., Rump, S. & Gladyshev, V. N. Comparative Genomics and Evolution of Molybdenum Utilization. *Coord. Chem. Rev.* **255**, 1206–1217 (2011).
- Pakharukova, M. Y. et al. Cytochrome P450 in fluke *Opisthorchis felineus*: identification and characterization. *Mol. Biochem. Parasitol.* **181**, 190–194 (2012).
- Kuntz, A. N. et al. Thioredoxin glutathione reductase from *Schistosoma mansoni*: an essential parasite enzyme and a key drug target. *PLoS Med.* **4**, e206 (2007).
- Zhong, Y. F. & Holland, P. W. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* **13**, 567–568 (2011).
- Holland, P. W. Beyond the Hox: how widespread is homeobox gene clustering? *J. Anat.* **199**, 13–23 (2001).
- Brooke, N. M., Garcia-Fernandez, J. & Holland, P. W. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**, 920–922 (1998).
- Riddiford, N. & Olson, P. D. *Wnt* gene loss in flatworms. *Dev. Genes Evol.* **221**, 187–197 (2011).
- Brehm, K. *Echinococcus multilocularis* as an experimental model in stem cell research and molecular host-parasite interaction. *Parasitology* **137**, 537–555 (2010).
- Parkinson, J. et al. A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl. Trop. Dis.* **6**, e1897 (2012).
- Raz, E. The function and regulation of vasa-like genes in germ-cell development. *Genome Biol.* **1**, R1017.1–R1017.6 (2000).
- Garcia-España, A. et al. Appearance of new tetraspanin genes during vertebrate evolution. *Genomics* **91**, 326–334 (2008).
- Díaz, A. et al. Understanding the laminated layer of larval *Echinococcus* I: structure. *Trends Parasitol.* **27**, 204–213 (2011).
- Hancock, K. et al. Characterization and cloning of GP50, a *Taenia solium* antigen diagnostic for cysticercosis. *Mol. Biochem. Parasitol.* **133**, 115–124 (2004).
- Heath, D. D., Jensen, O. & Lightowers, M. W. Progress in control of hydatidosis using vaccination - a review of formulation and delivery of the vaccine and recommendations for practical use in control programmes. *Acta Trop.* **85**, 133–143 (2003).
- Zhang, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
- Subjeck, J. R. & Repasky, E. A. Heat shock proteins and cancer therapy: the trail grows hotter! *Oncotarget* **2**, 433–434 (2011).
- Vargas-Parada, L., Solis, C. F. & Laclette, J. P. Heat shock and stress response of *Taenia solium* and *T. crassiceps* (Cestoda). *Parasitology* **122**, 583–588 (2001).
- Hemphill, A. et al. *Echinococcus* metacestodes as laboratory models for the screening of drugs against cestodes and trematodes. *Parasitology* **137**, 569–587 (2010).
- Brunetti, E. & White, A. C., Jr. Cestode infestations: hydatid disease and cysticercosis. *Infect. Dis. Clin. North Am.* **26**, 421–435 (2012).
- McVeigh, P. et al. Parasite neuropeptide biology: Seeding rational drug target selection? *Int. J. Parasitol. Drugs and Drug Res.* **2**, 76–91 (2012).
- Marks, N. J. & Maule, A. G. Neuropeptides in helminths: occurrence and distribution. *Adv. Exp. Med. Biol.* **692**, 49–77 (2010).
- Van Nassauw, L., Toovey, S., Van Op den Bosch, J., Timmermans, J. P. & Vercruyse, J. Schistosomicidal activity of the antimalarial drug, mefloquine, in *Schistosoma mansoni*-infected mice. *Travel Med. Infect. Dis.* **6**, 253–258 (2008).

44. Doudican, N., Rodriguez, A., Osman, I. & Orlow, S. J. Mebendazole induces apoptosis via Bcl-2 inactivation in chemoresistant melanoma cells. *Mol. Cancer Res.* **6**, 1308–1315 (2008).
45. Grove, D. I. *A History of Human Helminthology*. 848 (CAB International, 1990).
46. Spiliotis, M. & Brehm, K. Axenic *in vitro* cultivation of *Echinococcus multilocularis* metacestode vesicles and the generation of primary cell cultures. *Methods Mol. Biol.* **470**, 245–262 (2009).
47. Spiliotis, M. *et al.* *Echinococcus multilocularis* primary cells: improved isolation, small-scale cultivation and RNA interference. *Mol. Biochem. Parasitol.* **174**, 83–87 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Dunn and OpGen for the *E. multilocularis* optical map; Roche for 20kb 454 libraries; R. Rance, M. Quail, D. Willey, N. Smerdon and K. Oliver for sequencing libraries at WTSI; T. D. Otto for bioinformatics expertise; R. Davies and Q. Lin for help with data release; A. Bateman, M. Punta, P. Coggill and J. Minstry of Pfam; N. Rowlings from MEROPS database, J. Gough at SUPERFAMILY and C. Dessimoz for OMA; C. Seed and F. Jarero for laboratory assistance; J. Overington and B. Al-Lazikani; and B. Brejová for predicting genes with ExonHunter. P.D.O. and N.P.S. were supported in part by a BBSRC grant (BBG0038151) to P.D.O. P.D.O. and Ma.Z. were supported by a SynTax joint UK Research Council grant. The European Research Council supported P.W.H.H. and J.Paps. J.Parkinson and S.S.H. were supported by an operating grant from the Canadian Institute for Health Research (CIHR MOP#84556). G.S. was supported by FIRCA-NIH (grant TW008588) and the Universidad de la República, CSIC (grant CSIC 625). The *E. multilocularis*, *E. granulosus* and *H. microstoma* genome projects were funded by the Wellcome Trust through their core support of the Wellcome Trust Sanger Institute (grant 098051). K.B. was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG; BR2045/4-1). The *Taenia solium* Genome Project (IMPULSA 03) was supported by the Universidad Nacional Autónoma de México. The *Taenia solium* Genome Consortium thanks P. de la Torre, J. Yáñez, P. Gaytán, S. Juárez and J. L. Fernández for technical support; L. Herrera-Estrella and LANGEBIO, CINVESTAV-Irapuato and C. B. Shoemaker for sequencing support and other advice; and J. Watanabe (deceased), S. Sugano and Y. Suzuki for the construction and sequencing of the full-length cDNA library.

**Author Contributions** I.J.T., Ma.Z., N.H. and M.B. Wrote the manuscript; M.B., K.B., J.P.L., X.S. and X.C. conceived and designed the project; N.H., M.B. and Ma.Z. coordinated the project. R.J.B., P.D., C.F., T.H., J.H., K.L., X.L., S.H., N.M., P.D.O., M.R., E.S., N.P.S., Y.Z. and K.B. prepared parasite material and nucleic acids; I.J.T., Ma.Z., A.G., K.E. and A.S.F. were involved in genome assembly; I.J.T., K.L.B., A.T., H.B., S.N., T.H., A.G. and K.E. were involved in genome assembly improvement; I.J.T., Ma.Z., A.S.F., X.S., K.L., J.H., A.G. and K.E. were involved in gene predictions; I.J.T., Ma.Z., K.L.B., A.T., H.B., O.L., S.N., R.C., R.J.B., G.F., E.S., X.S., J.P.L., K.L., J.H., A.G., K.E., S.H. and X.C. were involved in gene annotation; N.D.S., M.A., J.A.K., K.E., J.H., S.H., X.S. and Y.Z. were involved in data processing, computational and bioinformatics support; I.J.T. analysed genome structure, comparative genomics and ploidy; A.S.F., I.J.T. and Ma.Z. analysed gene structure; T.H. and H.M.B. experimentally validated micro-exons; A.G., K.B., F.K. and I.J.T. were involved with *trans*-splicing and polycistrons; I.J.T., S.S.H., J.Parkinson, G.S. and Ma.Z. examined metabolism and detoxification; P.W.H.H., J.Paps, N.R. and P.D.O. examined homeobox gene loss; K.B., I.J.T. and Ma.Z. examined stem cell specializations; Ma.Z. and J.D.W.

examined domains; I.J.T., Ma.Z., C.F. and G.S. examined tapeworm-specific genes and expansions; Ma.Z. examined kinases and proteases; T.A.D. and Mo.Z. examined GPCRs; Mo.Z., T.A.D., U.K. and K.B. examined neuropeptides; M.R., L.K., F.C. and M.C. examined neuronal signalling; Ma.Z. examined drug targets; and G.S., M.R., C.F., K.B., P.W.H.H., P.D.O., A.G., R.J.B., G.F., E.S., X.S., J.P.L. and J.C. commented on the manuscript drafts.

**Author Information** The tapeworm genome projects were registered under the INSDC project IDs PRJEB122 (*E. multilocularis*), PRJEB121 (*E. granulosus*), PRJEB124 (*H. microstoma*) and PRJNA16816 (*T. solium*, Mexico). Illumina and 454 data are released to the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/>) under accession numbers ERP000351, ERP000452 and PRJNA16816. Capillary data is at <http://www.ncbi.nlm.nih.gov/Traces/trace.cgi>, SEQ\_LIB\_ID 98488, 98489 and 101760, CENTER NAME SC (*E. multilocularis*) and sg1, sg2, sg3, sg4 and sg5 (*T. solium*, Mexico). Genome data are available from <http://www.sanger.ac.uk/resources/downloads/helminths/> (*E. multilocularis*, *E. granulosus* and *H. microstoma*) and <http://www.teniasolium.unam.mx/taenia/> (*T. solium*). The complete genome annotation is available at <http://www.genedb.org>. All RNA-seq data were released to ArrayExpress under accession numbers E-ERAD-50 or E-ERAD-56. *T. solium* EST sequences were released to <http://www.ncbi.nlm.nih.gov/nucest/>, under accession numbers EL740221 to EL763490. Reprints and permissions information is available at <http://www.nature.com/reprints>. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (mb4@sanger.ac.uk), K.B. (kbrehm@hygiene.uni-wuerzburg.de) or J.P.L. (laclette@biomedicas.unam.mx).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

#### The *Taenia solium* Genome Consortium

Alejandro Garciarrubio<sup>1</sup>, Raúl J. Bobes<sup>2</sup>, Gladys Fragoso<sup>2</sup>, Alejandro Sánchez-Flores<sup>1</sup>, Karel Estrada<sup>1</sup>, Miguel A. Cevallos<sup>3</sup>, Enrique Morett<sup>4</sup>, Víctor González<sup>3</sup>, Tobias Portillo<sup>1</sup>, Adrian Ochoa-Leyva<sup>4</sup>, Marco V. José<sup>5</sup>, Edda Sciuotto<sup>2</sup>, Abraham Landa<sup>5</sup>, Lucía Jiménez<sup>5</sup>, Víctor Valdés<sup>6</sup>, Julio C. Carrero<sup>2</sup>, Carlos Larralde<sup>2</sup>, Jorge Morales-Montor<sup>2</sup>, Jorge Limón-Lason<sup>2</sup>, Xavier Soberón<sup>1,4</sup> & Juan P. Laclette<sup>2</sup>

<sup>1</sup>Institute of Biotechnology, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. <sup>2</sup>Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México, D.F. Mexico. <sup>3</sup>Genomic Sciences Center, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, Mexico.

<sup>4</sup>Instituto Nacional de Medicina Genómica, Periférico Sur No. 4809 Col. Arenal Tepepan, Delegación Tlalpan, 14610 México, D.F. México. <sup>5</sup>School of Medicine, Universidad Nacional Autónoma de México, 04510 México, D.F. Mexico. <sup>6</sup>School of Sciences, Universidad Nacional Autónoma de México, 04510 México, D.F. Mexico.

# SCF<sup>FBXL3</sup> ubiquitin ligase targets cryptochromes at their cofactor pocket

Weiman Xing<sup>1,2</sup>, Luca Busino<sup>3</sup>, Thomas R. Hinds<sup>1</sup>, Samuel T. Marionni<sup>4</sup>, Nabiha H. Saifee<sup>1</sup>, Matthew F. Bush<sup>4</sup>, Michele Pagano<sup>3,5</sup> & Ning Zheng<sup>1,2</sup>

The cryptochrome (CRY) flavoproteins act as blue-light receptors in plants and insects, but perform light-independent functions at the core of the mammalian circadian clock. To drive clock oscillations, mammalian CRYs associate with the Period proteins (PERs) and together inhibit the transcription of their own genes. The SCF<sup>FBXL3</sup> ubiquitin ligase complex controls this negative feedback loop by promoting CRY ubiquitination and degradation. However, the molecular mechanisms of their interactions and the functional role of flavin adenine dinucleotide (FAD) binding in CRYs remain poorly understood. Here we report crystal structures of mammalian CRY2 in its apo, FAD-bound and FBXL3-SKP1-complexed forms. Distinct from other cryptochromes of known structures, mammalian CRY2 binds FAD dynamically with an open cofactor pocket. Notably, the F-box protein FBXL3 captures CRY2 by simultaneously occupying its FAD-binding pocket with a conserved carboxy-terminal tail and burying its PER-binding interface. This novel F-box-protein-substrate bipartite interaction is susceptible to disruption by both FAD and PERs, suggesting a new avenue for pharmacological targeting of the complex and a multifaceted regulatory mechanism of CRY ubiquitination.

Cryptochromes (CRYs) are evolutionarily conserved FAD-binding proteins that share close sequence homology with DNA photolyases and are widespread in both plant and animal kingdoms<sup>1,2</sup>. Using FAD as a chromophore, plant cryptochromes function as blue-light photo-receptors and regulate a broad range of light responses during plant growth and development<sup>1,3</sup>. In animals, cryptochromes are classified into two types<sup>4,5</sup>. Type I cryptochromes, exemplified by *Drosophila melanogaster* CRY, maintain blue-light sensitivity and have a key role in photic entrainment of the insect circadian clock<sup>6</sup>, whereas type II cryptochromes are photo-insensitive and evolved as central components of the molecular clock<sup>7,8</sup>.

Mammalian CRYs, the prototypic type II cryptochromes, act together with PERs, CLOCK and BMAL1 (brain and muscle aryl hydrocarbon receptor nuclear translocator-like protein 1) to constitute a transcription negative feedback loop that oscillates with a circadian periodicity of approximately 24 h (ref. 9). In this molecular clockwork, CLOCK and BMAL1 heterodimerize and activate the transcription of CRY and PER genes, whose protein products in turn form complexes in the cytoplasm and translocate into the nucleus to suppress their own gene expression by binding and inhibiting CLOCK–BMAL1 (refs 10, 11). In mammals, this basic clock circuitry drives the circadian oscillations of both the master clock in the brain and the peripheral clocks throughout the body<sup>9,12</sup>. The master clock is synchronized to light–dark cycles perceived by the retina and coordinates clocks in the peripheral tissues through a variety of mechanisms, such as feeding behaviour, body temperature and hormonal secretion<sup>13–15</sup>.

Despite their functional divergence, many plant and animal cryptochromes are functionally linked to protein ubiquitination and degradation. For example, light-activated *Arabidopsis thaliana* CRY1 mediates seedling photomorphogenesis by inhibiting COP1 ubiquitin ligase<sup>16,17</sup>, whereas *Drosophila* CRY modulates the insect circadian

rhythm by promoting light-dependent degradation of the clock protein Timeless (TIM) and itself via the Jetlag ubiquitin ligase<sup>18,19</sup>. In both cases, photo-excitation of FAD is thought to induce conformational changes within the cryptochrome proteins and alter their interactions with the target proteins<sup>20,21</sup>. Recent biochemical and genetic studies have shown that mammalian CRYs are also functionally connected to the ubiquitin–proteasome system<sup>22–24</sup>. The SCF<sup>FBXL3</sup> ubiquitin ligase complex (SKP1–CUL1–F-box protein complex containing the F-box protein FBXL3) has been identified to regulate the clock by promoting the degradation of CRYs.

As an integral part of the clock circuitry, SCF<sup>FBXL3</sup>-catalysed CRY degradation needs to be regulated tightly. Indeed, the AMP-activated protein kinase (AMPK), a cellular metabolic sensor, has been shown to phosphorylate mouse CRY1 and accelerate its degradation by enhancing CRY1 binding to FBXL3 (ref. 25). Moreover, a recent circadian chemical screen has identified a period-lengthening small molecule that can compete with FAD for CRY binding and stabilize CRYs by inhibiting their ubiquitination<sup>26</sup>. These activities of the compound suggest that SCF<sup>FBXL3</sup>-mediated CRY ubiquitination may be tuneable by the FAD cofactor, whose function in mammalian CRYs remains elusive<sup>2</sup>. To provide the missing structural framework for understanding the functions of mammalian CRYs and the regulatory mechanisms of CRY–FBXL3 interaction, we report here the crystal structures of mammalian CRY2 in three different functional forms.

## Structure of murine apo CRY2 PHR

Cryptochromes in general contain a conserved photolyase-homology region (PHR) and a unique cryptochrome C-terminal (CCT) extension (CCE) (Supplementary Fig. 1a). We first determined the 2.7 Å crystal structure of the proteolytically stable mouse CRY2 PHR domain (amino acids 1–512) (Supplementary Fig. 2 and Supplementary Table 1). As expected, murine CRY PHR adopts a canonical photolyase

<sup>1</sup>Department of Pharmacology, Box 357280, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

<sup>3</sup>Department of Pathology, NYU Cancer Institute, New York University School of Medicine, 522 First Avenue, SRB 1107, New York, New York 10016, USA. <sup>4</sup>Department of Chemistry, Box 351700, University of Washington, Seattle, Washington 98195, USA. <sup>5</sup>Howard Hughes Medical Institute, New York University School of Medicine, 522 First Avenue, SRB 1107, New York, New York 10016, USA.

fold, consisting of an amino-terminal  $\alpha/\beta$  domain (N-terminal domain containing both  $\alpha$ -helices and  $\beta$ -sheets), a C-terminal  $\alpha$ -helical domain, and a connecting linker sequence (Fig. 1a). At the sequence level, vertebrate CRYs are more homologous to metazoan (6-4) photolyases than plant or type I animal cryptochromes (Supplementary Table 2). After superposition, murine CRY2 PHR can be aligned with the most homologous structure of *Drosophila* (6-4) photolyase with a root mean squared deviation (r.m.s.d.) of 1.1 Å over 431 equivalent C $\alpha$  positions (Supplementary Fig. 3). However, unlike all known structures of (6-4) photolyases and cryptochromes<sup>27</sup>, the murine CRY2 PHR structure represents the apo form of the flavoprotein (Fig. 1a). We found no electron density of FAD at its expected cofactor-binding pocket, despite the well-defined and conserved FAD-binding cavity<sup>28</sup>. This feature is in agreement with previous reports showing that only a small fraction of mammalian CRYs contains FAD when isolated from mammalian cells or various heterologous systems<sup>2</sup>. Similar to *Drosophila* (6-4) photolyase, murine CRY2 PHR does not co-purify with a second cofactor that acts as a light-harvesting antenna in some bacterial photolyases<sup>20,28–31</sup>.

### Structure of FAD-bound murine CRY2 PHR

Oxidized FAD is naturally fluorescent with an emission maximum at approximately 525 nm. In the presence of excess murine CRY2 PHR, FAD fluorescence is strongly quenched, suggesting that murine CRY2 PHR still retains specific FAD-binding activity (Fig. 1b). By titrating the amount of protein added, we estimated that FAD binds to isolated CRY2 PHR with an apparent dissociation constant ( $K_d$ ) of approximately 40  $\mu$ M (Fig. 1b). We next crystallized murine CRY2 PHR together with FAD and obtained a 2.2 Å resolution structure of the holoenzyme (Supplementary Table 1). As anticipated, FAD docks into the predicted cofactor pocket of murine CRY2 PHR and adopts the characteristic U-shaped conformation as seen in other cryptochrome and photolyase structures (Fig. 1c). Superposition analysis shows that its binding mode is nearly identical to that revealed in the *Drosophila* (6-4) photolyase structure (Fig. 1d).

A closer structural comparison, nevertheless, reveals a marked difference at the cofactor-binding site between murine CRY2 PHR and

other family members. In plant and insect (6-4) photolyases and cryptochromes, FAD is largely sequestered and buried deep inside the pocket as a prosthetic group<sup>20,28–30</sup>. In marked contrast, the cofactor is only partially embedded in murine CRY2 PHR, with one side of its adenosine diphosphate group fully exposed to the solvent (Fig. 2a, b). This unusual exposure of the cofactor can be attributed to the unique features of two adjacent structural elements in murine CRY2 PHR, the phosphate loop (a surface loop previously named for its phosphate-binding activity) and the protrusion motif. In *Drosophila* and *Arabidopsis* (6-4) photolyases, the adenine moiety of FAD is mostly hidden under the well-ordered phosphate loop<sup>28,30</sup>. This loop harbours a central lysine residue that hydrogen bonds with the FAD adenine N7 atom and is secured further by a nearby sequence known as the protrusion motif (Fig. 2c). In murine CRY2 PHR, the protrusion motif is shifted far away from FAD, whereas the phosphate loop becomes completely disordered in the crystal (Fig. 2c). These local structural variations in CRY2 result in an overall open conformation of the FAD-binding pocket, which might explain the moderate affinity of FAD.

### Overall structure of FBXL3–SKP1–CRY2

The unique CCEs of vertebrate CRYs share a strictly conserved 11-amino-acid segment (conserved CCE sequence, CCS), which is followed by a highly variable C-terminal region (Supplementary Fig. 1b). Using a protein co-expression strategy, we noticed that the nearly full-length form of murine CRY2 (amino acids 1–544) with the intact CCS motif can form a stable complex with FBXL3–SKP1. This prompted us to crystallize and determine a 2.8 Å resolution structure of the heterotrimeric complex (Supplementary Table 1).

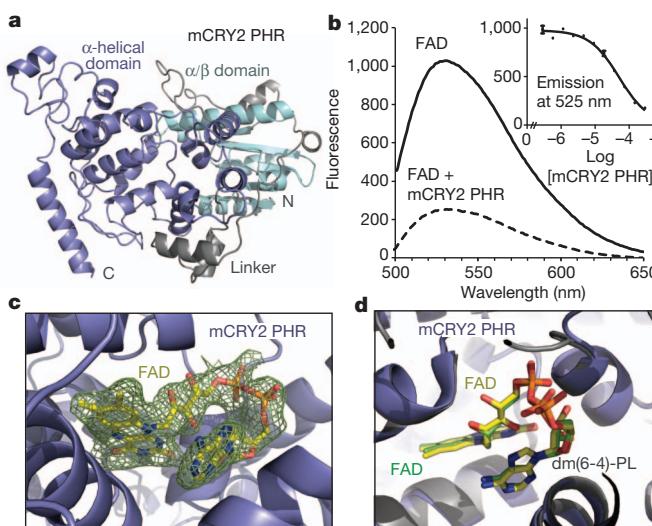
The overall architecture of the CRY2–FBXL3–SKP1 complex resembles an ice-cream cone with the globular CRY2 sitting on top of the cup-shaped FBXL3–SKP1 complex (Fig. 3). The FBXL3 protein harbours a canonical 3-helix F-box motif that interacts with the SCF adaptor protein SKP1, and its C-terminal leucine-rich-repeat (LRR) domain folds into a curved and sickle-shaped solenoid structure, whose concave surface wraps around the  $\alpha$ -helical domain and the CCS region of CRY2 opposite to the  $\alpha/\beta$  domain (Fig. 3 and Supplementary Fig. 4). The FBXL3–CRY2 complex buries more than 4,800 Å<sup>2</sup> solvent accessible surface area and spontaneously assembles in a phosphorylation-independent manner, which was confirmed by native protein mass spectrometry (Supplementary Fig. 5).

### The FBXL3 LRR domain

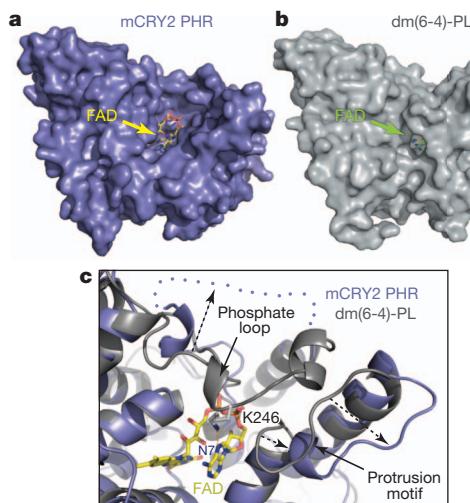
The FBXL3 LRR domain contains 12 LRRs that pack in tandem to produce a semicircular arch (Fig. 4a). The solenoid topology of the LRRs dictates that the concave surface of the arch is formed by the parallel  $\beta$ -strands of the repeats and the convex side is decorated by the  $\alpha$ -helices. We refer to the topside of the arch that is lined with intra-repeat loops as the apical ridge (Fig. 4b).

Although the FBXL3 LRRs fold into a continuous single domain, it can be separated into two obvious halves as suggested by a prominent structural irregularity in LRR7 and LRR8. Compared to the six N-terminal LRRs, LRR7 has a much longer  $\beta$ -strand, which causes a significant offset at the apical ridge (Fig. 4b). This feature continues in LRR8, but is lost in LRR9 to LRR12. With a similar offset at the basal side between LRR8 and LRR9, all four C-terminal LRRs seem to be shifted up, relative to the six N-terminal repeats. Overall, LRR7 and LRR8 mediate this spatial displacement and this hints at a possible gene fusion event when FBXL3 evolved in vertebrates. We name the two halves of FBXL3 LRRs as LRR-N (LRR1–LRR6) and LRR-C (LRR7–LRR12).

A quick inspection of the FBXL3 LRR–CRY2 interface reveals that the two halves of the FBXL LRR domain have unequal roles in engaging the cryptochrome protein. All six repeats in LRR-C are in close contact with the  $\alpha$ -helical domain of CRY2, whereas most parts of LRR-N, except its four apical loops, are separated from CRY2 by a gap (Fig. 4c). Consistent with a more important role of LRR-C in recruiting CRY2,



**Figure 1 | Structure of murine CRY2 PHR in apo- and FAD-bound forms.** **a**, Overall structure of the apo-bound form of the mouse CRY2 PHR (mCRY2 PHR) domain. **b**, mCRY2 PHR induced quenching of FAD fluorescence. The inset shows FAD-fluorescence quenching with mCRY2 PHR concentration titrated. Error bars, s.d. **c**, A close-up view of FAD bound to mCRY2 PHR with positive  $F_o - F_c$  electron density contoured at 1.5  $\sigma$  and calculated before the cofactor was built (green mesh). **d**, mCRY2 PHR and the *Drosophila* (6-4) photolyase (dm(6-4)-PL) are superimposed, with their FAD cofactors shown as sticks and their FAD carbon atoms in yellow and green, respectively.



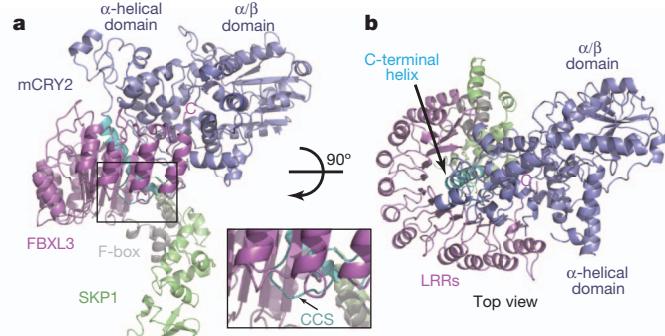
**Figure 2 | The open FAD-binding pocket of murine CRY2 PHR.** **a, b,** Surface representations of mCRY2 PHR and *Drosophila* (6-4) photolyase with their associated FAD cofactors shown as sticks. **c,** A close-up view of the open FAD-binding pocket of mCRY2-PHR shown with FAD as sticks and *Drosophila* (6-4)-photolyase superimposed (ribbons). Dotted blue line, disordered mCRY2 phosphate loop; dashed arrows, the differences of the phosphate loop and protrusion motif in the two proteins, respectively.

the missense mutations in the alleles known as after hours and overtime, Cys358Ser and Ile364Thr substitutions, are mapped to LRR10 and LRR11, respectively (Fig. 4c and Supplementary Fig. 6). Because both residues are located in the hydrophobic core of the LRR fold, their substitution by polar amino acids probably perturbs the local conformation of LRR-C.

### The C-terminal tail of FBXL3

The C-terminal tail of FBXL3 represents the hallmark of the FBXL3–CRY2 interface by sticking out from the flat concave surface of LRR-C and penetrating into CRY2 (Fig. 5a). The 12-amino-acid C-terminal region of FBXL3 is mostly invariant in vertebrates and terminates with a tryptophan residue (Supplementary Fig. 6). This sequence caps the LRR solenoid and then takes a sharp turn at Pro 422, inserting the 5 amino acids closest to the C terminus into the  $\alpha$ -helical domain of CRY2 (Supplementary Fig. 7). Notably, the last residue of the tail, Trp 428, reaches the core of the FAD-binding pocket and physically occupies the cofactor site.

Superposition analysis shows that the FBXL3 C-terminal tail approaches the FAD-binding pocket from the same angle as the (6-4) DNA lesion approaches the FAD-binding pocket in the *Drosophila*



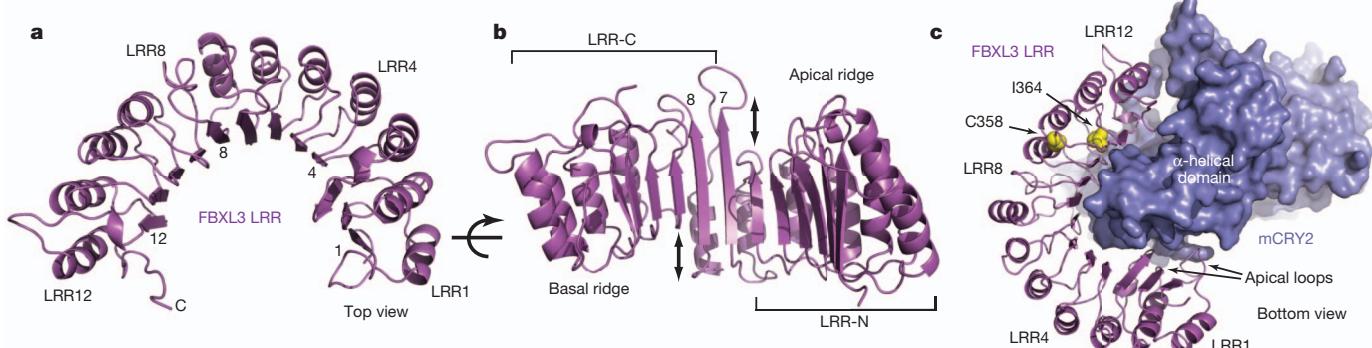
**Figure 3 | Overall structure of the CRY2–FBXL3–SKP1 complex.** **a,** Ribbon diagrams of the complex containing mCRY2 (blue), FBXL3 (magenta) and SKP1 (green). C, carboxyl terminus of FBXL3. The C-terminal helix and the CCS region in mCRY2 are shown in cyan and the F-box domain of FBXL3 is in grey. The inset shows a close-up view of the mCRY2 CCS region. **b,** Orthogonal view of the CRY2–FBXL3–SKP1 complex as shown in **a**.

(6-4) photolyase–DNA complex structure<sup>30</sup> (Fig. 5b). To expand the entrance, Met 425 of FBXL3 pushes aside an important CRY2 loop, which we name the interface loop. To enter the pocket, the FBXL3 tail stacks its Pro 426 residue against Trp 310 of CRY2 and flips the side chain of CRY2 His 373 with Thr 427 (Fig. 5c, d). In the FAD-bound form, these CRY2 residues directly contact the cofactor. Inside the pocket, the side chain of FBXL3 Trp 428 occupies the central space of the cavity, which corresponds to the narrow gap between the isoalloxazine and adenine rings of FAD. While donating a hydrogen bond to the backbone carbonyl group of Gln 307 in CRY2, the indole ring of Trp 428 makes multiple hydrophobic contacts with CRY2. The carboxyl group of Trp 428 forms additional hydrogen bonds with CRY2 Ser 414 and a signature arginine residue, Arg 376 (Fig. 5c, d). Overall, this network of interactions anchors the FBXL3 tail at the very centre of the FAD-binding pocket, predicting that FAD and the F-box protein will compete to bind the cryptochromes protein.

### Interface between FBXL3 and CRY2 PHR

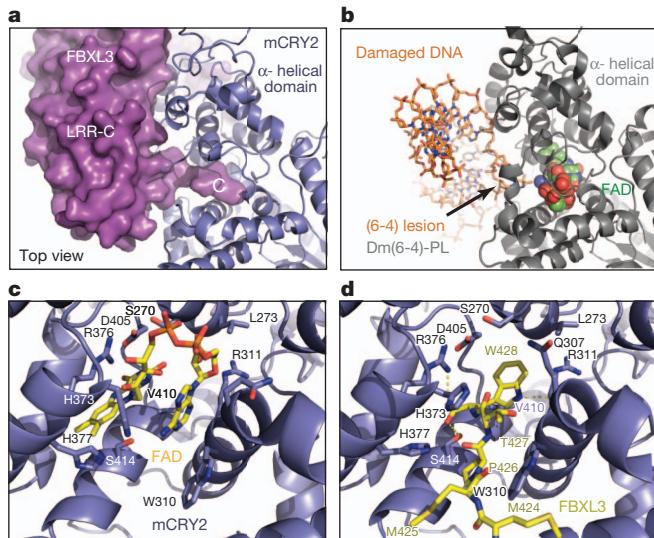
In addition to the C-terminal tail, the FBXL3 LRR domain also forms extensive interfaces with CRY2 through its solenoid fold. At the apical ridge of LRR-N, four FBXL3 loops from LRR1–LRR4 interact with the back of the  $\alpha$ -helical domain of CRY2 opposite to the FAD-binding pocket (Figs 3b and 4c). This part of the interface is mediated largely by polar interactions (Supplementary Fig. 7).

In contrast to LRR-N, the FBXL3 LRR-C subdomain binds CRY2 through a predominantly hydrophobic interface. On its concave wall, LRR-C presents a cluster of more than ten hydrophobic residues strictly



**Figure 4 | Structure of the FBXL3 LRR domain.** **a,** Ribbon diagrams of the LRR domain of FBXL3 with its complete C-terminal tail. Select LRRs are labelled and numbered at their helices and  $\beta$ -strands. **b,** An orthogonal view of the LRR domain shown in **a**. Double-headed arrows indicate the offset between

LRR6 and LRR7, and between LRR8 and LRR9. **c,** A bottom view of the FBXL3 LRR domain (magenta) bound to mCRY2 (blue). Side chains of the FBXL3 residues mutated in the overtime and after hours alleles are shown as yellow spheres.

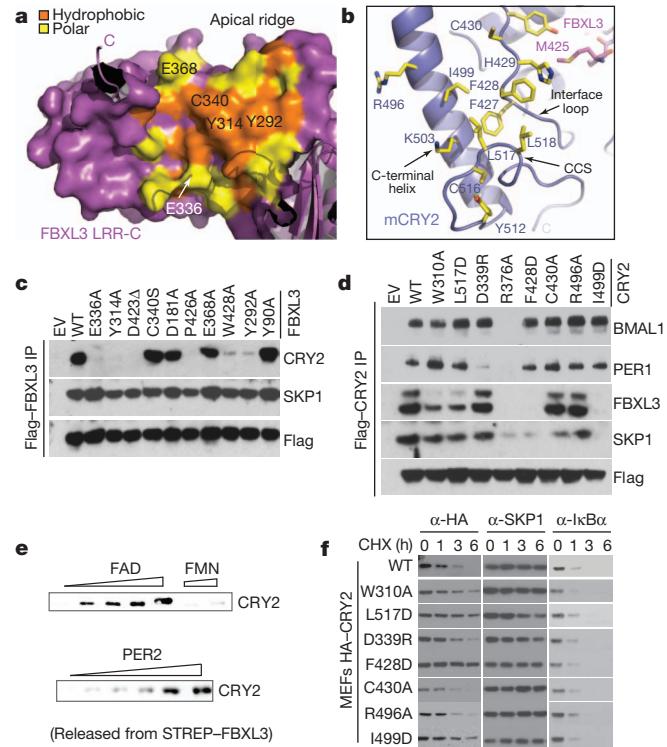


**Figure 5 | Interaction between the FBXL3 C-terminal tail and the CRY2 FAD-binding pocket.** **a**, Surface representation of the FBXL3 LRR-C subdomain with its C-terminal tail inserted into the FAD-binding pocket of mCRY2 (blue ribbons). **b**, The structure of *Drosophila* (6-4) photolyase–DNA complex shown from the same orientation as mCRY2 in **a**. FAD is shown as spheres. The damaged DNA substrate is shown as a stick diagram. **c, d**, A close-up view of the mCRY2 FAD-binding pocket showing key FAD- and FBXL3-interacting residues. FAD and the FBXL3 tail are shown as sticks. Dashed yellow lines represent hydrogen bonds.

conserved in vertebrate CRYs (Fig. 6a and Supplementary Fig. 6). These residues, plus a circle of peripheral polar amino acids, recognize a large complementary surface on CRY2 adjacent to the cofactor pocket. The interface on the CRY2 side is constructed by three key structural motifs; the interface loop, the C-terminal helix and the CCS region (Fig. 6b). With a string of strictly conserved aromatic residues (Supplementary Fig. 8), the CRY2 interface loop is buttressed by Met 425 from the FBXL3 tail on one side and the CRY2 C-terminal helix on the other. The CRY2 CCS region adopts a well-ordered loop structure that wraps back onto the interface loop and the C-terminal helix. Together, these three motifs contribute a panel of hydrophobic residues grouped at the centre of the LRR-C binding interface. By making direct contacts with both FBXL3 and CRY2 PHR, the conserved CCS region of CRY2 probably takes up its structural role at the interface through an ‘induced fit’ mechanism (Supplementary Fig. 9).

To map regions that are critical for FBXL3–CRY2 interaction, we tested the binding activity of a series of FBXL3 and CRY2 mutants. Disruption of interactions at the apical ridge of LRR-N did not affect the complex formation (Asp181Ala and Tyr90Ala (Fig. 6c), Asp339Arg and Arg496Ala (Fig. 6d and Supplementary Fig. 7)), indicating a supportive role of LRR-N in recruiting CRY2. By contrast, CRY2 binding was impaired severely by either single amino acid substitution or truncation of the FBXL3 tail (Pro426Ala, Trp428Ala, Asp423Δ (Fig. 6c)). Similarly, alteration of a CRY2 residue in the cofactor pocket also compromised complex association (Trp310Ala (Fig. 5d, 6d)). These results pinpoint the FBXL3 C-terminal tail as a ‘hot spot’ of the interface and suggest a possible antagonistic role of FAD in regulating the FBXL3–CRY2 interaction. In a purified system, we confirmed that free FAD, but not flavin mononucleotide (FMN), was sufficient to disrupt a pre-formed FBXL3–CRY2 complex in a dose-dependent manner (Fig. 6e).

Our mutagenesis analysis also established the FBXL3 LRR-C subdomain as a second critical docking site for CRY2. Mutations of individual aromatic residues at the hydrophobic concave wall of LRR-C or a nearby CCS-interacting residue effectively abolished CRY2 binding (Tyr292Ala, Tyr314Ala and Glu336Ala (Fig. 6a, c and Supplementary Fig. 9)). The same effect was seen when a single



**Figure 6 | Structural and functional analyses of the FBXL3–murine CRY2 interface.** **a**, Surface and ribbon diagrams of the FBXL3 LRR-C subdomain showing hydrophobic and polar residues that are involved in mCRY2 binding. Amino acids selected for mutational analyses are labelled. The C-terminal tail and some N-terminal repeats are shown as ribbon diagrams. **b**, Ribbon diagrams of the mCRY2 structural elements involved in FBXL3 LRR-C interaction. Select interface residues are shown as sticks. A part of the FBXL3 tail (magenta) is shown for reference. **c**, Interactions of retrovirus-expressed FBXL3 mutants with endogenous mCRY2 in mouse embryonic fibroblasts (MEF) assessed by co-immunoprecipitation (IP) and western blot analysis. **d**, Flag-mCRY2 complexes were immunoprecipitated with an anti-Flag resin from transfected HEK293T cells and assessed for binding to FBXL3, PER1, BMAL1 and SKP1. Arg376Ala served as a negative control as it is probably detrimental to mCRY2 folding (see Supplementary Discussion for Cys430Ala and FBXL3–Cys340Ser mutants). **e**, Two *in vitro* competition assays showing the ability of FAD and PER2, but not FMN, to disrupt a preformed FBXL3–mCRY2 complex. **f**, MEFs were infected with retroviruses expressing either mCRY2 or mCRY2 mutants and treated with cycloheximide (CHX) for the indicated time in hours. Total extracts were examined by immunoblot analysis.

hydrophobic-to-charged mutation was introduced to each of the three LRR-C-interacting structural motifs in CRY2; that is, the interface loop (Phe428Asp), the C-terminal helix (Ile499Asp) and the CCS region (Leu517Asp) (Fig. 6b, d). The functional importance of this part of the interface and the FBXL3 tail-gripping pocket was underscored further by the stabilization of the CRY2 mutants with hot-spot residues altered at these sites (Fig. 6f and Supplementary Fig. 9).

The C-terminal helix of CRY2 and its two basic residues, Arg 501 and Lys 503, are thought to be essential for PER binding<sup>32,33</sup>. In complex with FBXL3, the entire CRY2 C-terminal helix is masked by the LRR domain of the F-box protein (Fig. 3b and Supplementary Fig. 9). We found by chance that mutation of Asp 339, which is in close proximity to Arg 501, also showed defective interaction with PER1 (Fig. 6d and Supplementary Fig. 7). These lines of evidence strongly suggest that PERs and FBXL3 share an overlapping CRY2-binding interface, which should prevent them from binding CRY2 simultaneously. Indeed, PER2-bound CRY2 is completely devoid of FBXL3 (Supplementary Fig. 10). Importantly, just like FAD, a purified C-terminal fragment of PER2, which corresponds to a previously mapped minimal CRY-binding region<sup>34</sup>, was able to dislodge CRY2 from FBXL3 (Fig. 6e). We were able

to confirm that the abundance of PER2 had a direct effect on the stability of both CRY1 and CRY2 (Supplementary Fig. 10)<sup>34,35</sup>, probably by competing with the SCF<sup>FBXL3</sup> E3 ligase.

## Discussion

The structure of the mammalian cryptochrome reveals an unexpected open FAD-binding pocket that probably evolved in vertebrates to enable dynamic FAD binding. This unique property is built on the plasticity of the phosphate loop guarding the entrance of the pocket. Intriguingly, this loop is highly conserved among vertebrates (Supplementary Fig. 8), raising the possibility that its alteration by phosphorylation<sup>25,36</sup> or CRY-binding proteins might regulate the affinity of FAD (Supplementary Fig. 11).

Importantly, our structural results have unravelled the functional requirement for the dynamic FAD-binding pocket in CRYs and a novel mechanism of substrate-recognition by the SCF ubiquitin ligase. In contrast to the canonical degron-based substrate-engagement scheme<sup>37,38</sup>, FBXL3 takes advantage of the deep but accessible cofactor pocket of CRYs to capture the substrate (Supplementary Fig. 12). This interaction mode establishes FAD and its analogues as potential pharmacological agents for controlling clock oscillation by competing with SCF<sup>FBXL3</sup>. Such an effect might be achieved by the CRY-stabilizing compounds identified in the recent circadian chemical screen<sup>26</sup>.

The unexpected interface between CRY2 and FBXL3 suggests that their interactions can be regulated by multiple mechanisms. As a common metabolic redox cofactor in the cell, it is possible that FAD directly controls the stability of CRYs by competing with the ubiquitin ligase. Although oxidized FAD binds murine CRY2 with a modest affinity (see Supplementary Discussion), its interaction with the CRY proteins can be influenced by its redox states and several factors mentioned above. The CRY2–FBXL3 structure also supports the idea that PERs may have a role in stabilizing CRYs by shielding them from SCF<sup>FBXL3</sup>. This function is reminiscent of the stabilization of PERs by CRYs against SCF<sup>βTrCP</sup>-mediated degradation<sup>11,34</sup>. Such interplay between the two obligate functional partners could serve as a mechanism for synchronizing their stability to ensure robust clock oscillation and phase shift. In fact, the CRY1 destabilizing effect of Ser 71 phosphorylation by AMPK may be partially explained by its reported activity in blocking PER2 binding<sup>25</sup>.

## METHODS SUMMARY

The full Methods provides detailed information about all experimental procedures, including description of protein purification; protein crystallization, data collection and structure determination; FAD fluorescence assay; *in vitro* competition assay; and mutagenesis, cell culture, binding and stability analyses.

**Full Methods** and any associated references are available in the online version of the paper.

Received 21 August 2012; accepted 29 January 2013.

Published online 17 March 2013.

- Chaves, I. et al. The cryptochromes: blue light photoreceptors in plants and animals. *Annu. Rev. Plant Biol.* **62**, 335–364 (2011).
- Ozturk, N. et al. Structure and function of animal cryptochromes. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 119–131 (2007).
- Lin, C. & Shalitin, D. Cryptochrome structure and signal transduction. *Annu. Rev. Plant Biol.* **54**, 469–496 (2003).
- Yuan, Q., Metterville, D., Briscoe, A. D. & Reppert, S. M. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. *Mol. Biol. Evol.* **24**, 948–955 (2007).
- Zhu, H. et al. The two CRYs of the butterfly. *Curr. Biol.* **15**, R953–R954 (2005); erratum **16**, 730 (2006).
- Stanewsky, R. et al. The *cry<sup>b</sup>* mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell* **95**, 681–692 (1998).
- van der Horst, G. T. et al. Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398**, 627–630 (1999).
- Griffin, E. A., Staknis, D. & Weitz, C. J. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* **286**, 768–771 (1999).
- Reppert, S. M. & Weaver, D. R. Coordination of circadian timing in mammals. *Nature* **418**, 935–941 (2002).

- Shearman, L. P. et al. Interacting molecular loops in the mammalian circadian clock. *Science* **288**, 1013–1019 (2000).
- Lee, C., Etchegaray, J. P., Cagampang, F. R., Loudon, A. S. & Reppert, S. M. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107**, 855–867 (2001).
- Dibner, C., Schibler, U. & Albrecht, U. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu. Rev. Physiol.* **72**, 517–549 (2010).
- Green, C. B., Takahashi, J. S. & Bass, J. The meter of metabolism. *Cell* **134**, 728–742 (2008).
- Bass, J. & Takahashi, J. S. Circadian integration of metabolism and energetics. *Science* **330**, 1349–1354 (2010).
- Asher, G. & Schibler, U. Crosstalk between components of circadian and metabolic cycles in mammals. *Cell Metab.* **13**, 125–137 (2011).
- Yang, H. Q., Tang, R. H. & Cashmore, A. R. The signaling mechanism of *Arabidopsis* CRY1 involves direct interaction with COP1. *Plant Cell* **13**, 2573–2587 (2001).
- Wang, H., Ma, L. G., Li, J. M., Zhao, H. Y. & Deng, X. W. Direct interaction of *Arabidopsis* cryptochromes with COP1 in light control development. *Science* **294**, 154–158 (2001).
- Peschel, N., Chen, K. F., Szabo, G. & Stanewsky, R. Light-dependent interactions between the *Drosophila* circadian clock factors Cryptochrome, Jetlag, and Timeless. *Curr. Biol.* **19**, 241–247 (2009).
- Koh, K., Zheng, X. & Sehgal, A. JETLAG resets the *Drosophila* circadian clock by promoting light-induced degradation of TIMELESS. *Science* **312**, 1809–1812 (2006).
- Zoltowski, B. D. et al. Structure of full-length *Drosophila* cryptochrome. *Nature* **480**, 396–399 (2011).
- Liu, B., Liu, H., Zhong, D. & Lin, C. Searching for a photocycle of the cryptochrome photoreceptors. *Curr. Opin. Plant Biol.* **13**, 578–586 (2010).
- Busino, L. et al. SCF<sup>FBXL3</sup> controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *Science* **316**, 900–904 (2007).
- Godinho, S. I. et al. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science* **316**, 897–900 (2007).
- Siepka, S. M. et al. Circadian mutant *Overtime* reveals F-box protein FBXL3 regulation of Cryptochrome and *Period* gene expression. *Cell* **129**, 1011–1023 (2007).
- Lamia, K. A. et al. AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science* **326**, 437–440 (2009).
- Hirota, T. et al. Identification of small molecule activators of cryptochrome. *Science* **337**, 1094–1097 (2012).
- Müller, M. & Carell, T. Structural biology of DNA photolyases and cryptochromes. *Curr. Opin. Struct. Biol.* **19**, 277–285 (2009).
- Hitomi, K. et al. Functional motifs in the (6–4) photolyase crystal structure make a comparative framework for DNA repair photolyases and clock cryptochromes. *Proc. Natl Acad. Sci. USA* **106**, 6962–6967 (2009).
- Brautigam, C. A. et al. Structure of the photolyase-like domain of cryptochrome 1 from *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **101**, 12142–12147 (2004).
- Maul, M. J. et al. Crystal structure and mechanism of a DNA (6–4) photolyase. *Angew. Chem. Int. Edn Engl.* **47**, 10076–10080 (2008).
- Park, H. W., Kim, S. T., Sancar, A. & Deisenhofer, J. Crystal structure of DNA photolyase from *Escherichia coli*. *Science* **268**, 1866–1872 (1995).
- Ozber, N. et al. Identification of two amino acids in the C-terminal domain of mouse CRY2 essential for PER2 interaction. *BMC Mol. Biol.* **11**, 69 (2010).
- Chaves, I. et al. Functional evolution of the photolyase/cryptochrome protein family: importance of the C terminus of mammalian CRY1 for circadian core oscillator performance. *Mol. Cell. Biol.* **26**, 1743–1753 (2006).
- Yagita, K. et al. Nucleocytoplasmic shuttling and mCRY-dependent inhibition of ubiquitylation of the mPER2 clock protein. *EMBO J.* **21**, 1301–1314 (2002).
- Chen, R. et al. Rhythmic PER abundance defines a critical nodal point for negative feedback within the circadian clock mechanism. *Mol. Cell* **36**, 417–430 (2009).
- Sanada, K., Harada, Y., Sakai, M., Todo, T. & Fukada, Y. Serine phosphorylation of mCRY1 and mCRY2 by mitogen-activated protein kinase. *Genes Cells* **9**, 697–708 (2004).
- Hao, B. et al. Structural basis of the Cks1-dependent recognition of p27<sup>Kip1</sup> by the SCF<sup>Skp2</sup> ubiquitin ligase. *Mol. Cell* **20**, 9–19 (2005).
- Tan, X. et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–645 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the beamline staff of the Advanced Light Source at the University of California at Berkeley for help with data collection, and members of the Zheng laboratory for discussion. This work is supported by the Howard Hughes Medical Institute (N.Z. and M.P.), the National Institutes of Health (R01-CA107134 to N.Z., 5T32-HL007151 to L.B., and R01-GM057587, R37-CA-076584 and R21-CA161108 to M.P.), and the University of Washington (S.T.M. and M.F.B.).

**Author Contributions** The protein purification and crystallization experiments were conceived by W.X., L.B., M.P. and N.Z., initiated by N.H.S., and conducted by W.X., W.X. and N.Z. determined and analysed the structures. FAD fluorescence and *in vitro* competition experiments were conceived by W.X., T.R.H. and N.Z., and conducted by W.X. and T.R.H. Mutational and binding studies, and stability analyses were conceived by L.B., M.P., W.X. and N.Z., and conducted by L.B., S.T.M. and M.F.B. conducted native mass spectrometry experiments.

**Author Information** Structural coordinates and structural factors for FBXL3–CRY2–SKP1, CRY2–FAD and CRY2 are deposited in the Protein Data Bank under accession numbers 4I6J, 4I6G and 4I6E. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Z. (nzheng@u.washington.edu).

## METHODS

**Recombinant protein purification.** The mouse CRY2 (amino acids 1–544) was expressed as a glutathione S-transferase (GST) fusion protein in High Five (Invitrogen) suspension insect cells and isolated by glutathione affinity chromatography using buffer containing 50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 10% Glycerol, 10 mM DTT (dithiothreitol) and 0.5 mM FAD. The protein was purified further by cation exchange and gel filtration chromatography after on-column cleavage by tobacco etch virus (TEV) protease. During purification, the polypeptide produced a series of C-terminal degradation products that were not affected by the presence or absence of FAD and dithionite. Two shorter versions of the protein, murine CRY2 with amino acids 1–527 and murine CRY2 with amino acids 1–512, were subsequently purified. Native mass spectrometry analysis confirmed that murine CRY2 with amino acids 1–512 was stable, but murine CRY2 with amino acids 1–527 was not (Supplementary Figs 1 and 2). The mouse CRY2 (amino acids 1–544) protein was co-expressed with full-length human FBXL3 and SKP1 in High Five suspension insect cells (human FBXL3 is 97% identical to mouse FBXL3, and the F-box and LRR regions of human and mouse FBXL3 differ by only 3 amino acids; Supplementary Fig. 6). FBXL3 was fused with an N-terminal GST tag, whereas the other two proteins were tag-free. The CRY2–FBXL3–SKP1 complex was isolated from the soluble cell lysate by glutathione affinity chromatography. After on-column tag cleavage by TEV, the complex was purified further by anion exchange and gel filtration chromatography and then concentrated by ultrafiltration to 10 mg ml<sup>-1</sup> in a buffer of 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM DTT. Mouse PER2 (amino acids 1095–1235) was overexpressed as a GST-fusion protein in *Escherichia coli* and purified by glutathione affinity, anion exchange and gel filtration chromatography.

**Crystallization, data collection and structure determination.** The crystals of the CRY2–FBXL3–SKP1 complex were grown at 4 °C by the hanging-drop vapour diffusion method, using 1.5 µl protein complex sample mixed with an equal volume of reservoir solution containing 0.2 M ammonium citrate, 13–14% PEG3350 and 7% acetonitrile. Diffraction-quality crystals that were obtained were subjected to a post-crystallization dehydration procedure by gradually increasing the concentration of PEG3350 to 30%, and then directly frozen in liquid nitrogen. This procedure consistently improved the resolution of the crystals beyond approximately 4 Å. The CRY2–FBXL3–SKP1 derivative crystals were prepared by soaking the native crystals in the buffer containing 15% PEG3350 and 7% acetonitrile supplemented with 0.2 mM KAu(CN)<sub>2</sub> for 12 h, followed by soaking in 0.5 mM KAu(CN)<sub>2</sub> for 12 h and then 1 mM KAu(CN)<sub>2</sub> for a further 12 h. After soaking, the crystals were dehydrated by the same method as native crystals. The crystals of murine CRY2 were also grown at 4 °C by the hanging-drop vapour diffusion method with 1.5 µl protein mixed with an equal volume of reservoir solution containing 0.1 M MES-imidazole, pH 6.5, 7.2% PEG8K, 20% ethylene glycol (v/v), 0.03 M NaF, NaBr and NaI. The crystals were directly frozen in liquid nitrogen. The CRY2–FAD crystals were obtained by soaking the murine CRY2 crystals with 1 mM FAD in the reservoir buffer with extra 7.5% Glycerol for 2 h before freezing in liquid nitrogen. All data sets were collected at the BL8.2.1 and BL8.2.2 beamlines at the Advanced Light Source of the Lawrence Berkeley National Laboratory. Reflection data were indexed, integrated and scaled with the HKL2000 package<sup>39</sup>. The murine CRY2 PHR domain structure was determined by molecular replacement using the *Drosophila* (6-4) photolyase structure (PDB:3CVU) as the search model. Soaking with FAD altered the unit cell and substantially improved the murine CRY2 PHR domain crystals. For the apo structure, translation libration screw-motion (TLS) refinement was necessary to reduce the *R*-free value below 0.30. The CRY2–FBXL3–SKP1 structure was solved by molecular replacement combined with single-wavelength anomalous dispersion phases (MR-SAD) using the software PHENIX<sup>40</sup>. The structural models were manually built, refined and rebuilt with the programs COOT<sup>41</sup> and PHENIX<sup>40</sup>.

**FAD fluorescence assay.** FAD fluorescence and binding studies were carried out in a PBS buffer, pH 7.4, with 10% glycerol at room temperature in a half-area 96-well plate with a total volume of 100 µl. Interaction was quantified by measuring the decrease in FAD fluorescence (excitation at 450 nm and emission at 520 nm) after binding to mouse CRY2. FAD concentration was held constant at 10 µM in all wells, and murine CRY2 PHR was serially diluted ten times from 300 µM to 292 nM. Nonspecific FAD binding (background) was measured in parallel with equal molar concentrations of ovalbumin or lysozyme, which do not bind FAD. All protein concentrations were run in duplicate. The binding data was fit using a standard nonlinear regression curve fitting method with a log(inhibitor) versus response model in Prism 5 (GraphPad Software).

**In vitro competition assay.** Strep-tagged full-length FBXL3, GST-tagged murine CRY2 (amino acids 1–544) and His-tagged SKP1 were co-expressed in monolayer High Five insect cells. The complex was purified without elution by Strep-Tactin affinity resin in a buffer containing 100 mM Tris-HCl, pH 8.0, 300 mM NaCl and 20 mM DTT. The Strep-Tactin resin beads with the immobilized complex were subsequently aliquoted in a volume of 100 µl with 1 mg ml<sup>-1</sup> total protein and placed in a panel of 7.5 mm gravity columns. The beads were incubated with 300 µl of oxidized FAD, FMN or purified murine PER2 (amino acids 1095–1235) in a serial dilution (FAD: 0, 5, 50, 500 µM and 5 mM; FMN: 50 and 500 µM; murine PER2: 0, 2.5, 5, 10, 20 and 40 µM) for 3 h at 4 °C. Flow-through and subsequent two-column-volume wash fractions were collected from each column and the GST–CRY2 protein released from the column was examined with SDS-PAGE and detected by western blot analysis using anti-GST antibody.

**Cell culture.** HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% bovine serum. Immortalized mouse embryonic fibroblasts (MEFs) were maintained in DMEM containing 10% fetal bovine serum (FBS). HEK293T cells were transfected with PEI (Polysciences) according to the manufacturer's instructions. Cycloheximide (Sigma) was used at a final concentration of 100 µM. For retrovirus production, GP-293 packaging cells (Clontech) were used. Forty-eight hours after transfection, the virus-containing medium was collected and supplemented with 8 µg ml<sup>-1</sup> polybrene (Sigma). Cells were then infected by replacing the cell culture medium with the viral supernatant for six hours. Forty-eight hours after infection, cells were selected with 1 µg ml<sup>-1</sup> puromycin.

**Co-immunoprecipitation binding studies.** Cells were collected and subsequently lysed in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 50 mM NaF, 0.5% NP40, plus protease and phosphatase inhibitors). Immunoprecipitation was carried out with anti-Flag agarose beads (Sigma). Immuno-complexes were washed five times with lysis buffer and eluted in Laemmli buffer containing 1% SDS. Cell extracts or Flag-immunocomplexes were analysed by SDS-PAGE and then by immunoblot analysis with the following antibodies: anti-Flag (Sigma), anti-CRY1 and CRY2 (Bethyl), anti-PER1 (Alpha Diagnostic), anti-PER2 (Bethyl), anti-BMAL1 (Bethyl), anti-Skp1 (Invitrogen) and anti-FBXL3 (ref. 22).

**Plasmids.** Wild-type and mutant mouse CRY2 complementary DNAs were subcloned into pcDNA3.1-Flag. Mouse PER2 was cloned into pMT-Myc. For retrovirus production, cDNAs encoding Flag or HA-tagged FBXL3 or CRY2 wild-type and mutants, were subcloned into the retroviral vector pBabePuro (Cell Biolabs). Point mutations were generated using the QuikChange Site-directed Mutagenesis kit (Stratagene).

39. Otwinowski, Z. & Minor, W. In *Methods in Enzymology* Vol. 276 (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic Press, 1997).
40. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
41. Collaborative Computational Project, number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

# Femtosecond switching of magnetism via strongly correlated spin-charge quantum excitations

Tianqi Li<sup>1,2</sup>, Aaron Patz<sup>1,2</sup>, Leonidas Mouchliadis<sup>3,4</sup>, Jiaqiang Yan<sup>2†</sup>, Thomas A. Lograsso<sup>2</sup>, Ilias E. Perakis<sup>3,4</sup> & Jigang Wang<sup>1,2</sup>

The technological demand to push the gigahertz ( $10^9$  hertz) switching speed limit of today's magnetic memory and logic devices into the terahertz ( $10^{12}$  hertz) regime underlies the entire field of spin-electronics and integrated multi-functional devices. This challenge is met by all-optical magnetic switching based on coherent spin manipulation<sup>1</sup>. By analogy to femtosecond chemistry and photosynthetic dynamics<sup>2</sup>—in which photoproducts of chemical and biochemical reactions can be influenced by creating suitable superpositions of molecular states—femtosecond-laser-excited coherence between electronic states can switch magnetic order by ‘suddenly’ breaking the delicate balance between competing phases of correlated materials: for example, manganites exhibiting colossal magneto-resistance suitable for applications<sup>3,4</sup>. Here we show femtosecond ( $10^{-15}$  seconds) photo-induced switching from antiferromagnetic to ferromagnetic ordering in  $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ , by observing the establishment (within about 120 femtoseconds) of a huge temperature-dependent magnetization with photo-excitation threshold behaviour absent in the optical reflectivity. The development of ferromagnetic correlations during the femtosecond laser pulse reveals an initial quantum coherent regime of magnetism, distinguished from the picosecond ( $10^{-12}$  seconds) lattice-heating regime characterized by phase separation without threshold behaviour<sup>5,6</sup>. Our simulations reproduce the nonlinear femtosecond spin generation and underpin fast quantum spin-flip fluctuations correlated with coherent superpositions of electronic states to initiate local ferromagnetic correlations. These results merge two fields, femtosecond magnetism in metals and band insulators<sup>1,7–9</sup>, and non-equilibrium phase transitions of strongly correlated electrons<sup>10–17</sup>, in which local interactions exceeding the kinetic energy produce a complex balance of competing orders.

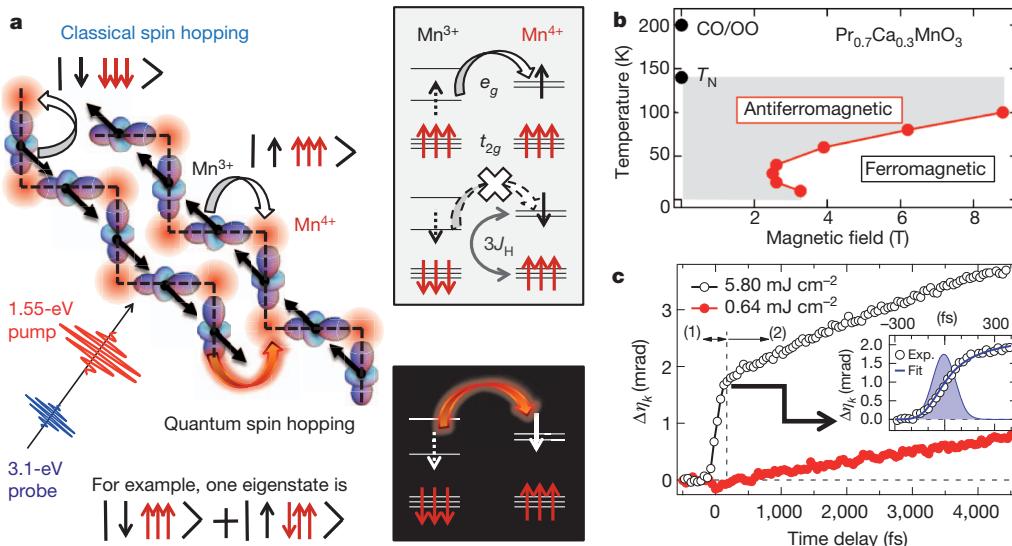
There is growing evidence that femtosecond transient polarization of condensed matter systems during a laser pulse can be used to manipulate spin or change magnetic order<sup>1,7–9,18–20</sup>. Non-adiabatic coherent photo-excitation during very early times can be used to control subsequent slower dynamics driven by the free energy or excited potential surface<sup>21</sup>. In strongly correlated materials, such as colossal magneto-resistive (CMR) manganite and high- $T_c$  copper oxide systems, prior ultrafast spectroscopy experiments mostly concentrated on laser-induced melting of electronic charge order and orbital order (CO/OO), and on structural phase transformations<sup>15–17</sup>. Thus far experiments directly probing magnetism only showed picosecond and nanosecond spin generation, much slower than the femtosecond excitation pulse, without threshold dependence on laser intensity<sup>5,6,22,23</sup>. At such ‘long timescales’, lattice heating and phase separation dynamics dominate, as seen, for example, in some CMR manganites<sup>5,6</sup> (see also ref. 24, where it is shown that laser heating of materials with high temperature ferromagnetic orders can induce magnetization). Transient magnetism there was mostly indirectly inferred by transferring static pictures and quasi-equilibrium theories to the ultrafast regime.

Here we excite the antiferromagnetic (AFM) CO/OO ground state of CMR manganites shown in Fig. 1a, which consists of one-dimensional zig-zag chains with alternating manganese  $\text{Mn}^{4+}/\text{Mn}^{3+}$  atoms (charge order) and ferromagnetically-aligned spins. This particular antiferromagnetic structure is called CE-type AFM<sup>3,4</sup>. Neighbouring chains are AFM-ordered, which restricts inter-chain electron hopping because of the large magnetic energy penalty (Hund's rule). The  $\text{Mn}^{4+}$  atom  $t_{2g}$  orbitals<sup>3,4</sup> are occupied by three electrons forming local  $S = 3/2$  spins,  $S_i$ . In the  $\text{Mn}^{3+}$  atoms, an additional electron of spin  $s_i$  populates a  $\text{Mn} e_g$  orbital<sup>3,4</sup>. The resulting Jahn-Teller distortion of  $\text{MnO}_6$  octahedra lifts the  $e_g$  orbital degeneracy on the  $\text{Mn}^{3+}$  atoms, which are then populated by alternating  $d_{3x^2-r^2}/d_{3y^2-r^2}$  orbitals (orbital order)<sup>3,4</sup>. The measured phase diagram of our  $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$  manganite, Fig. 1b, results from the competition between these AFM CO/OO insulating and ferromagnetic metallic phases. The ground state is however always insulating under weak magnetic fields ( $B < 0.5$  T in our experiment), for all relevant temperatures. It exhibits CO/OO below  $\sim 200$  K and CE-type AFM below the Néel temperature,  $T_N \approx 140$  K (shaded area). Thus, heating of our system cannot induce ferromagnetism, unlike those systems with high-temperature ferromagnetic orders<sup>24</sup>.

This Letter reports a femtosecond AFM to ferromagnetic phase transition in  $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ . In particular, we observe a threshold excitation fluence for inducing a huge magnetization during coherent nonlinear optical excitation of the CE-type AFM CO/OO ground state by a 82-fs laser pulse. To see this, we performed two-colour—near-infrared pump (1.55 eV), ultraviolet probe (3.1 eV)—femtosecond-resolved magnetic circular dichroism (MCD) and magneto-optical Kerr rotation (MOKE) measurements. This geometry allows us to observe, at 3.1 eV, genuine femtosecond spin dynamics of the  $t_{2g}$  electron magnetization component  $S_z$  perpendicular to the sample surface, induced by the pump (see Methods). We measure femtosecond changes in the complex magneto-optical angle  $\hat{\theta}_k = \theta_k + i\eta_k$ , where the imaginary part  $\eta_k$  (MCD) and real part  $\theta_k$  (MOKE) correspond to absorption difference and phase shift, respectively, between right- and left-circularly polarized light, induced by  $S_z(t)$ .

Typical temporal profiles of the pump-induced ellipticity change,  $\Delta\eta_k$ , during the first 5 ps are shown in Fig. 1c, at 30 K and for two pump fluences.  $\Delta\eta_k > 0$  indicates photo-induced magnetization from the AFM ground state. At low pump fluence,  $0.64 \text{ mJ cm}^{-2}$  (red filled circles), we only observe a gradual, picosecond magnetization rise. An increase of pump fluence to  $5.8 \text{ mJ cm}^{-2}$  (open circles) reveals a remarkable femtosecond regime of spin photogeneration: a huge quasi-instantaneous jump in magnetization, three orders of magnitude larger than in magnetic semiconductors<sup>20,25</sup>, now precedes picosecond spin dynamics. Deconvolution of the  $\Delta\eta_k$  and probe pulse temporal profiles give a fast magnetization rise time of  $\sim 120$  fs (Fig. 1c, inset). The clear discontinuity of  $\Delta\eta_k$  (dashed line, Fig. 1c) marks two different temporal regimes of transient magnetism. Figure 2a shows the detailed pump-fluence dependence of  $\Delta\eta_k$ : it reveals a photo-excitation threshold,  $I_{\text{th}} \approx 2.5 \text{ mJ cm}^{-2}$ , for the

<sup>1</sup>Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA. <sup>2</sup>Ames Laboratory – USDOE, Ames, Iowa 50011, USA. <sup>3</sup>Department of Physics, University of Crete, Heraklion, Crete 71003, Greece. <sup>4</sup>Institute of Electronic Structure and Laser, Foundation for Research and Technology-Hellas, Heraklion, Crete 71110, Greece. <sup>†</sup>Present addresses: Department of Materials Science and Engineering, University of Tennessee, Knoxville, Tennessee 37996, USA; Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.



**Figure 1 | A quantum many-body scheme for femtosecond switching of magnetism and ultrafast photo-induced spin dynamics in  $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ .** **a**, Schematics of ultrafast excitations of the CE-type AFM/CO/OO order, whose spin, charge, and orbital lattice pattern, shown in the left panel, consists of one-dimensional AFM-coupled zig-zag chains. For classical spins (right top panel), electron conduction and optical transitions are restricted within the same one-dimensional ferromagnetic chain because  $J = S - 1/2$  total spin configurations are suppressed by Hund's rule (white arrow, classical spin hopping). Quantum spin-flip fluctuations (right bottom), however, allow  $e_g$  electrons to hop on sites with opposite local  $t_{2g}$  spin, by forming non-equilibrium total spin eigenstates (see equation (1)) as illustrated (red arrow, quantum spin hopping). **b**, The equilibrium phase diagram of our  $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$  sample as a function of temperature and magnetic field (sweeping up). The red dots show the measured

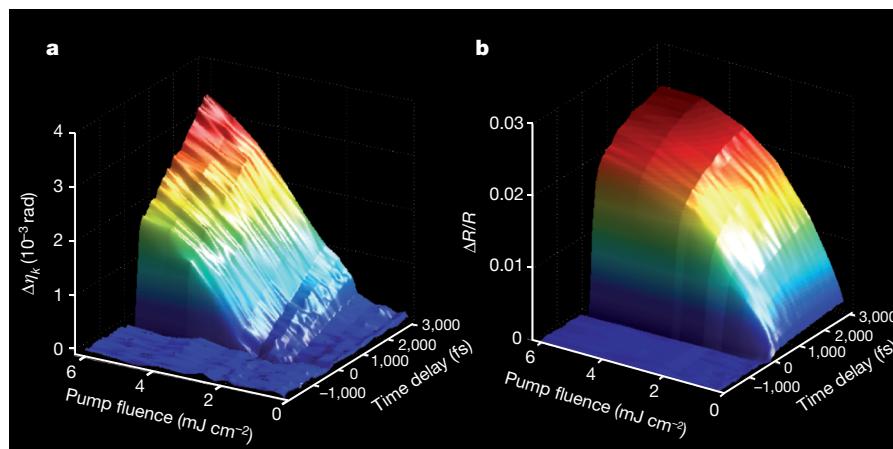
emergence of femtosecond magnetization. In contrast, the charge dynamics, measured by the pump-induced reflectivity change  $\Delta R/R$  (Fig. 2b), exhibits no threshold, which rules out melting of the CO/OO as the main origin of  $I_{\text{th}}$ . The critical fluence dependence exclusively for the femtosecond spin dynamics represents a hallmark of non-equilibrium magnetic phase transitions.

Further evidence for femtosecond change of magnetic ordering is presented in Fig. 3a, which compares the MCD  $\Delta\eta_k$  and MOKE  $\Delta\theta_k$  at two pump fluences and demonstrates notable differences between femtosecond and picosecond time intervals. Both MOKE and MCD demonstrate the existence of a fluence threshold for observing a distinct

boundary between the antiferromagnetic and ferromagnetic phases. Note that, for sufficiently small magnetic fields, the antiferromagnetic insulating phase persists at all temperatures. **c**, Ultrafast photo-induced ellipticity change  $\Delta\eta_k$  for magnetic field  $B = 0.5 \text{ T}$  and temperature  $T = 30 \text{ K}$  for low ( $0.64 \text{ mJ cm}^{-2}$ , red dots) and high ( $5.8 \text{ mJ cm}^{-2}$ , open circles) pump photo-excitation intensity. The femtosecond magnetism (temporal regime (1)) is clearly distinguished from the picosecond magnetism (temporal regime (2)) and only appears for high intensity. Inset, the instrument response time (shaded) and the convolution of the probe pulse with a double-exponential time-dependent magnetization increase. We extract time constants  $\tau_1 = 120 \text{ fs}$  (fast femtosecond magnetization) and  $\tau_2 = 18,500 \text{ fs}$  (slow picosecond magnetization). Open circles, experimental results (Exp); solid curve, double-exponential fit (Fit).

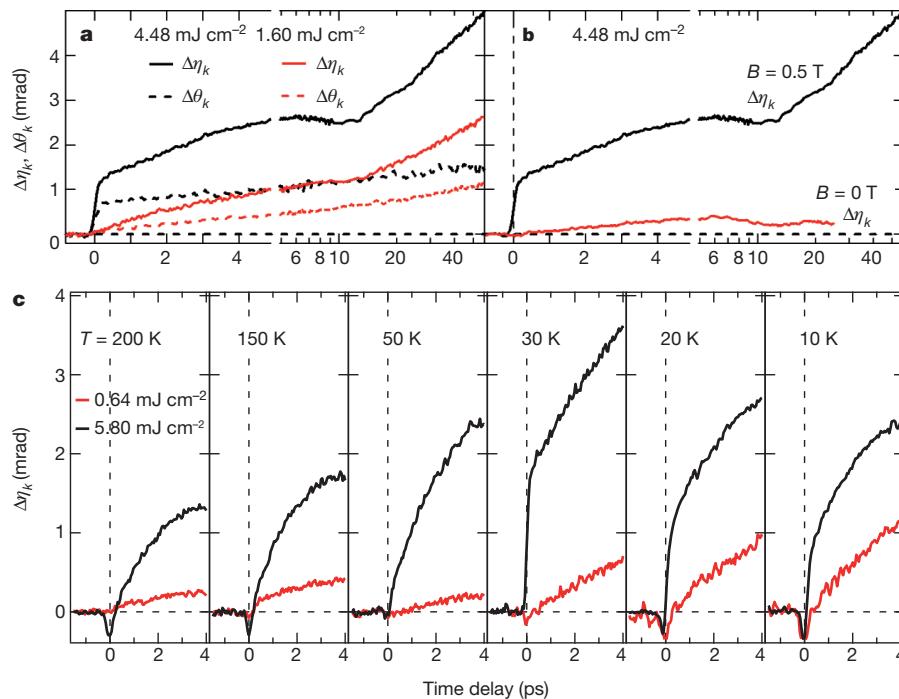
femtosecond pump-induced magnetization, with the same  $I_{\text{th}}$  in both (Supplementary Information). Such coincidence corroborates our conclusions. In contrast, the picosecond components of  $\Delta\eta_k$  and  $\Delta\theta_k$  exhibit different temporal profiles (for example, a damped oscillation is absent in MOKE) without threshold, indicative of slow thermal non-magnetic contributions<sup>5,6</sup>. In Fig. 3b, the femtosecond magnetization vanishes when the magnetic field is reduced from  $0.5 \text{ T}$  to  $0 \text{ T}$ , again pointing to the magnetic origin of the ‘discontinuity’ in  $\Delta\eta_k$  and  $\Delta\theta_k$  and, thereby, the fast spin generation during the excitation pulse.

Furthermore, the observed femtosecond spin photogeneration depends sensitively on temperature and AFM order (Fig. 3c). At



**Figure 2 | A three-dimensional view demonstrating the distinct femtosecond spin and charge dynamics as well as photo-excitation threshold behaviour.** **a**, Time-resolved ellipticity change  $\Delta\eta_k$  and **b**, differential optical reflectivity  $\Delta R/R$  as function of pump fluence. Measurements are taken under

the same experimental conditions. A photo-excitation threshold  $I_{\text{th}} \approx 2.5 \text{ mJ cm}^{-2}$  is seen for emergence of femtosecond magnetization ( $\Delta\eta_k$ ), whereas the charge dynamics exhibits no threshold. Magnetic field  $B = 0.5 \text{ T}$  and temperature  $T = 30 \text{ K}$ .



**Figure 3 | Evidence for genuine femtosecond switching of magnetic ordering and its sensitive temperature dependence.** **a**, Comparison between ellipticity  $\Delta\eta_k$  (solid lines) and MOKE  $\Delta\theta_k$  (dashed lines) under pump excitation fluences of  $4.48 \text{ mJ cm}^{-2}$  (black) and  $1.60 \text{ mJ cm}^{-2}$  (red). **b**, Magnetic field dependence of the ultrafast photo-induced ellipticity change during the first 5 ps (split axis) and for extended time intervals up to 60 ps

(logarithmic scale). Black,  $B = 0.5 \text{ T}$ ; red,  $B = 0 \text{ T}$ .  $T = 30 \text{ K}$ . **c**, Temperature dependence of the photo-induced ellipticity for the first 4 ps under two pump excitation fluences of  $0.64 \text{ mJ cm}^{-2}$  (red) and  $5.80 \text{ mJ cm}^{-2}$  (black). Note the non-monotonic dependence of the femtosecond photo-induced magnetization on temperature: it peaks at around 30 K.

200 K and 150 K, that is, above the AFM phase transition, we observe negligible femtosecond magnetization components for all fluences used. The femtosecond magnetization appears at 50 K above threshold, reaches its maximum around 30 K, and then decreases again at lower temperatures of 20 K and 10 K. We attribute this behaviour to the competition between spin stiffness—describing the rigidity of the AFM order—and thermal fluctuations, which display opposite temperature dependence. We note that the measured temperature for maximum femtosecond spin generation, 30 K, corresponds to the minimum critical  $B$  field required for driving an AFM to ferromagnetic phase transition in Fig. 1b.

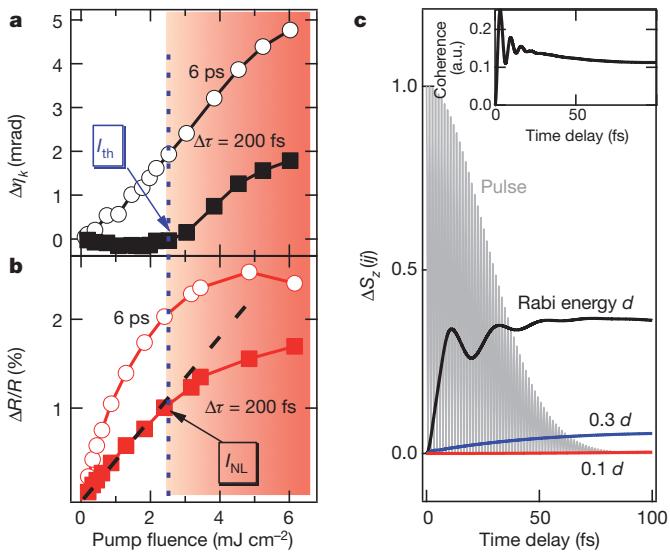
Figure 4a and b compares the pump-fluence dependence of  $\Delta\eta_k$  and  $\Delta R/R$  at two fixed times for the data shown in Fig. 2a and b,  $\Delta t = 200 \text{ fs}$  and  $6 \text{ ps}$ : we make two observations. First, any threshold dependence of photo-induced magnetization clearly smears out in the picosecond  $\Delta\eta_k$  data (black open circles), which shows a linear power dependence, and is absent in both picosecond (red open circles) and femtosecond (filled red squares)  $\Delta R/R$ . Thresholdless behaviours can be attributed to phonon-assisted phase separation<sup>5,6</sup>. However, the distinct non-linear pump fluence and threshold dependence of the femtosecond spin dynamics (Fig. 4a) reveal a new femtosecond magnetism non-thermal temporal regime during coherent nonlinear photo-excitation, well-distinguished from the picosecond thermal temporal regime. Second, the femtosecond build-up of ferromagnetic correlations above threshold  $I_{\text{th}}$  coincides with a transition, at  $I_{\text{NL}}$ , from linear to non-linear intensity dependence of  $\Delta R/R$  (shaded area). Such coincidence suggests that global electronic conducting paths emerging from photo-induced interchain hopping are important for femtosecond (but not picosecond) magnetism.

Can we attribute our observation of femtosecond magnetism to the photodoping of the AFM ground state with electrons and holes? A transition from AFM to ferromagnetic states with electron doping was predicted for Hamiltonians without CO/OO<sup>26</sup>, but these states are unstable to phase separation: ferromagnetic polarons, that is, an electron

surrounded by a ferromagnetically aligned neighbouring spin ‘cluster’, form locally inside the AFM matrix<sup>26</sup>. Above critical doping, polaron percolation results in a global ferromagnetic ground state. At equilibrium, electron doping of the CE/CO/OO state (Fig. 1a,  $x = \frac{1}{2} + \Delta x$ ) gives an unconventional ferromagnetic metal ( $x < 1/2$ ), whereas the system remains CE-AFM insulating with hole doping ( $x > 1/2$ ). Such electron–hole asymmetry, important for photodoping effects, arises from competition between electron delocalization (by canting anti-parallel  $t_{2g}$  spins) and self-trapping (by Jahn–Teller distortions)<sup>27</sup>. However, the latter lattice processes are slow (picosecond)<sup>23</sup>. During femtosecond times shorter than the Jahn–Teller phonon period and thermalization times, magnetization can only be induced by fast spin-flip/canting non-equilibrium processes.

Below, we propose a non-equilibrium spin-exchange mechanism that produces ferromagnetic correlations during the femtosecond laser pulse. Classical spin scenarios<sup>3,23</sup> neglect spin-flips, and so electron hopping from atom to atom diminishes with increasing angle between the local spins<sup>3</sup>. Consequently, interchain  $e_g$ -electron hopping is suppressed by AFM spin alignment, resulting in electronic confinement within one-dimensional chains (Fig. 1a, white arrows). This, however, suppresses any femtosecond spin dynamics, as the  $e_g$  and  $t_{2g}$  spins remain parallel within the same chain<sup>23</sup>. We must thus turn to quantum spin scenarios and, additionally, engage the transient coherence, correlation and nonlinearity that dominate during the laser coherent excitation, a new paradigm in ultrafast magnetism. Quantum spin-flips (Fig. 1a, red arrows) allow for electron hopping to a site with antiparallel local spin, as we can now populate the degenerate non-equilibrium quantum states of photo-excited and local electron spins shown schematically in Fig. 1a, which diagonalize the strong Hund’s rule interaction:

$$|i\alpha M\rangle = \sqrt{\frac{S+M+\frac{1}{2}}{2S+1}} c_{ia\uparrow}^\dagger |iS,M-\frac{1}{2}\rangle + \sqrt{\frac{S-M+\frac{1}{2}}{2S+1}} c_{ia\downarrow}^\dagger |iS,M+\frac{1}{2}\rangle \quad (1)$$



**Figure 4 | Interchain femtosecond quantum spin-flip fluctuations establish ferromagnetic correlations during the coherent laser photo-excitation.**

a, Pump fluence dependence of photo-induced  $\Delta\eta_K$  and b,  $\Delta R/R$ , after two time intervals,  $\Delta\tau = 200$  fs (filled squares) and  $\Delta\tau = 6$  ps (open circles), following photo-excitation, at 30 K. These summarize the data presented in Fig. 2a and b. The nonlinear fluence dependence with threshold  $I_{th}$  marked by the vertical dashed blue line is only seen for the femtosecond  $\Delta\eta_K$ . Note the different behaviour between the red shaded area (high intensity revealing femtosecond magnetism) and the low intensity regime (only picosecond magnetism). Although the photo-induced optical reflectivity change does not display any threshold behaviour, the high intensity red shaded regime above  $I_{NL}$  demonstrates a saturation nonlinear behaviour, while for low intensities below  $I_{NL}$  we observe the linear dependence on photo-excitation intensity expected from the usual perturbation expansion in terms of the optical field. c, Calculated photo-induced total spin for three Rabi energy values similar to those in the experiment,  $d = 70$  meV (black),  $0.3d$  (blue),  $0.1d$  (red), and inter-atomic coherence (inset) for  $d = 70$  meV. The laser pulse time-dependence is superimposed (grey;  $E^2(t)$ ), demonstrating that ferromagnetic local correlations transiently build-up from the AFM ground state (total spin zero) during the photo-excitation, with nonlinear dependence on Rabi energy and hence intensity. The time-dependence of the interatomic coherence describes ‘photo-induced bond’ order, which implies strong transient coupling between neighbouring atoms during the lifetime of the non-thermal populations, that is, before the establishment of a quasi-equilibrium hot temperature. The damped oscillations in the photo-induced spin, with period determined by the Jahn–Teller energy, reflect deviations from adiabaticity.

where  $|iSS_z\rangle$  are the  $Mn^{4+} t_{2g}$ -spin eigenstates. The eigenvalues of the total spin  $J_i = s_i + S_i$  are  $J = S + 1/2$  and  $M = -J, \dots, J$ .  $\hat{e}_{\alpha\sigma}^\dagger$  adds a spin- $\sigma$  electron in orbital  $\alpha$  at site  $i$ . The quantum-coherent superposition given in equation (1), suppressed in the classical spin limit  $S \rightarrow \infty$ , allows for ultrafast quantum spin dynamics, driven by the off-diagonal Hund’s rule magnetic interaction  $J_H S_i^\pm \cdot s_i^\mp$ .

In the insulating AFM state, we expect localized photo-excitations<sup>26</sup>. We thus describe the spin, charge and orbital populations at the  $i$ th atom by the diagonal density matrix elements  $\rho_i(m) = \langle iSm \rangle \langle Smi \rangle$  ( $Mn^{4+}$ ) and  $\rho_i(\alpha M) = \langle i\alpha M \rangle \langle \alpha M i \rangle$  ( $Mn^{3+}$ ). The 3.1-eV magneto-optical signal probes the  $t_{2g}$  spin<sup>28</sup>  $S_z(i) = \sum_m S \cos \theta_m \rho_i(m) + \sum_{\alpha M} S \cos \theta_M \rho_i(\alpha M)$ , where  $\cos \theta_m = m/S$  and  $\cos \theta_M = M/J$ . The 1.55-eV femtosecond pump field, tuned across the Jahn–Teller insulator gap, drives coherent electron hopping between neighbouring atoms  $i$  and  $j$  with different Jahn–Teller distortions. Such quantum kinetics is hidden in static measurements<sup>29</sup>, and the equations of motion couple  $\rho_i(m)$  and  $\rho_i(\alpha M)$  to the off-diagonal density matrix  $\langle \hat{e}_{\alpha'\sigma}^\dagger(jM') \hat{e}_{\alpha\sigma}(iM) \rangle$  describing time-dependent electron–hole superpositions of electronic quantum states given by equation (1) in two different atoms  $i$  and  $j$  (bond orders). The composite fermions created

by  $\hat{e}_{\alpha\sigma}^\dagger(iM) = |i\alpha M\rangle \left\langle iS, M - \frac{\sigma}{2} \right|$  treat the strong correlation (see Supplementary Information)<sup>30</sup>.

The solution of our density matrix equations for a two-site inter-chain cluster with ground state  $M = S + 1/2$  (Jahn–Teller-distorted site  $i$ ) and  $m = -S$  (empty site  $j$ ) numerically demonstrates the development of ferromagnetic correlations with femtosecond time-dependence and nonlinear dependence on photo-excitation similar to Fig. 1c, the key features of the femtosecond quantum spin regime. The calculated photo-induced spin  $\Delta S_z(ij) = \Delta S_z(i) + \Delta S_z(j)$  is shown in Fig. 4c for three Rabi energies, together with the laser pulse (here the Rabi energy is  $d = eEa$ , where  $E$  is the optical field,  $e$  is the electron charge, and  $a$  is the lattice spacing). The decrease in Rabi energy, from  $d = 70$  meV (black) to  $0.1d$  (red trace), diminishes the femtosecond spin dynamics as the perturbative regime is recovered for small  $d$ . Ferromagnetic correlation,  $\Delta S_z(ij) > 0$ , develops during laser excitation, driven by nonlinear photo-excited interatomic coherences  $\langle \hat{e}_{\alpha'\sigma}^\dagger(jM') \hat{e}_{\alpha\sigma}(iM) \rangle$  (see inset, Fig. 4c for  $d = 70$  meV) and electron–hole asymmetry. This spin correlation is enhanced by fast virtual electron hopping due to non-thermal/coherent nonlinear population changes and photo-induced bonding. The hole, created at  $Mn^{3+}$  site  $i$  populated by parallel  $e_g$  and  $t_{2g}$  spins, leaves the  $t_{2g}$  spin unchanged. The electron is created at the neighbouring  $Mn^{4+}$  site  $j$  populated by  $t_{2g}$  spin  $S_z = S$  (intrachain excitation) or  $S_z = -S$  (interchain excitation). It populates a  $Mn^{3+}$  quantum state of the form given by equation (1) with  $M = S_z + \sigma/2$  after interacting with the local spin. Interchain electron–hole excitations increase the  $S_z = -S$  spin via spin-flip (equation (1)), thus inducing ferromagnetic correlations between AFM-aligned chains. Intrachain excitations move the  $S + 1/2$  total spin along the chain and thus only affect  $\Delta R/R$  and the charge distribution. The collective behaviour of these photo-induced ferromagnetic ‘defects’ is affected by a small magnetic field, which creates a preferred direction through spin canting<sup>27</sup>, resulting in non-zero collective spin and enhanced interchain hopping. Such a  $B$ -field can also facilitate percolation of the photo-induced magnetic clusters, establishing macroscopic ferromagnetic order above a critical  $I_{th}$  as the local ferromagnetic cluster size grows nonlinearly. Our predicted femtosecond quantum spin canting/flip also gives rise to interchain conductive paths, which may explain  $I_{NL}$  in the femtosecond-resolved  $\Delta R/R$  (Fig. 4b).

## METHODS SUMMARY

**Sample growth and characterization.** Single crystals of  $Pr_{0.7}Ca_{0.3}MnO_3$  were grown by the floating-zone method in flowing oxygen with a growth rate of  $3\text{ mm h}^{-1}$ . X-ray powder diffraction on pulverized single crystals confirmed the single phase. The phase diagram in Fig. 1b was constructed by measuring the temperature- and field-dependence of magnetic susceptibility and electrical resistivity of our samples. Strong magnetic fields melt the AFM/CO/OO insulator ground state and induce a transition to a ferromagnetic metallic state, for example, at a critical magnetic field of  $B = 5\text{ T}$  at  $T \approx 75\text{ K}$ . This cannot be accessed by heating the sample. For our ultrafast measurements, we placed the sample in a weak magnetic field of no more than  $0.5\text{ T}$ , pointing perpendicular to the sample surface.

**Ultrafast magneto-optical spectroscopy.** A Ti:sapphire amplifier with centre wavelength of  $800\text{ nm}$  (1.55 eV), pulse duration of  $82\text{ fs}$  at the sample position, and  $1\text{ kHz}$  repetition rate was separated into pump and probe beams. The probe was frequency-doubled to  $400\text{ nm}$  (3.1 eV). This photon energy produces a large magneto-optical activity due to optical transitions between oxygen  $2p$  and  $Mn t_{2g}$  states<sup>28</sup>, and thereby mostly measures dynamics of  $t_{2g}$  local spins. The probe beam was kept nearly perpendicular to the sample surface (polar MOKE geometry) ( $\sim 7^\circ$  from the normal). To a first approximation, the magneto-optical signal is proportional to  $k \cdot S$ , where  $k$  is the vector of the probe light. Thus, the signal measured reflects the out-of-plane magnetization component  $S_z$  of  $t_{2g}$  local spins. Our pump beam is linearly polarized, with a tunable excitation fluence of up to  $\sim 6\text{ mJ cm}^{-2}$ . There is no obvious pump polarization dependence. This two-colour pump–probe geometry was shown to maximize the magneto-optical response while minimizing the contamination of magneto-optical signals during femtosecond timescales, for example, by dichroic bleaching, spectral weight transfer associated with the electronic phase transitions, and so on. More information on data analysis and theoretical modelling that solves the time-dependent equations of motion using the Runge–Kutta routines is given in Supplementary Information.

Received 25 September 2012; accepted 21 January 2013.

1. Bigot, J. Y., Vomir, M. & Beaurepaire, E. Coherent ultrafast magnetism induced by femtosecond laser pulses. *Nature Phys.* **5**, 515–520 (2009).
2. Brumer, P. & Shapiro, M. Controlling chemical reactions in the quantum regime. *Phys. Today* **64**, 11 (2011).
3. Dagotto, E., Hotta, T. & Moreo, A. Colossal magnetoresistant materials: the key role of phase separation. *Phys. Rep.* **344**, 1–153 (2001).
4. Tokura, Y. & Nagaosa, N. Orbital physics in transition-metal oxides. *Science* **288**, 462–468 (2000).
5. McGill, S. A. *et al.* Optical evidence for transient photoinduced magnetization in  $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ . *Phys. Rev. B* **71**, 075117 (2005).
6. Mertelj, T., Mamin, R., Yusupov, R. & Mihailovic, D. Ultrafast phase separation dynamics in  $\text{La}_{0.875}\text{Sr}_{0.125}\text{MnO}_3$  single crystals. *Phys. Rev. B* **83**, 113103 (2011).
7. Kirilyuk, A. V. *et al.* Ultrafast optical manipulation of magnetic order. *Rev. Mod. Phys.* **82**, 2731–2784 (2010).
8. Boeglin, C. *et al.* Distinguishing the ultrafast dynamics of spin and orbital moments in solids. *Nature* **465**, 458–461 (2010).
9. Kimel, A. V. *et al.* Ultrafast non-thermal control of magnetization by instantaneous photomagnetic pulses. *Nature* **435**, 655–657 (2005).
10. Liu, M. *et al.* Terahertz-field-induced insulator-to-metal transition in vanadium dioxide metamaterial. *Nature* **487**, 345–348 (2012).
11. Fiebig, M. *et al.* Sub-picosecond photo-induced melting of a charge-ordered state in a perovskite manganite. *Appl. Phys. B* **71**, 211–215 (2000).
12. Schmitt, F. *et al.* Transient electronic structure and melting of a charge density wave in  $\text{TbTe}_3$ . *Science* **321**, 1649–1652 (2008).
13. Fausti, D. *et al.* Light-induced superconductivity in a stripe-ordered cuprate. *Science* **331**, 189–191 (2011).
14. Kirn, K. W. *et al.* Ultrafast transient generation of spin-density-wave order in the normal state of  $\text{BaFe}_2\text{As}_2$  driven by coherent lattice vibrations. *Nature Mater.* **11**, 497–501 (2012).
15. Gedik, N., Yang, D., Logvenov, G., Bozovic, I. & Zewail, A. H. Nonequilibrium phase transitions in cuprates observed by ultrafast electron crystallography. *Science* **316**, 425–429 (2007).
16. Rini, M. *et al.* Control of the electronic phase of a manganite by mode-selective vibrational excitation. *Nature* **449**, 72–74 (2007).
17. Ichikawa, H. *et al.* Transient photoinduced hidden phase in a manganite. *Nature Mater.* **10**, 101–105 (2011).
18. Zhang, G. P. *et al.* Paradigm of the time-resolved magneto-optical Kerr effect for femtosecond magnetism. *Nature Phys.* **5**, 499–502 (2009).
19. Kapetanakis, M. D., Perakis, I. E., Wickey, K. J., Piermarocchi, C. & Wang, J. Femtosecond coherent control of spins in  $(\text{Ga}, \text{Mn})\text{As}$  ferromagnetic semiconductors using light. *Phys. Rev. Lett.* **103**, 047404 (2009).
20. Wang, J. *et al.* Memory effects in photoinduced femtosecond magnetization rotation in ferromagnetic  $\text{GaMnAs}$ . *Appl. Phys. Lett.* **94**, 021101 (2009).
21. Kapetanakis, M. D., Lingos, P. C., Piermarocchi, C., Wang, J. & Perakis, I. E. All-optical four-state magnetization reversal in  $(\text{Ga}, \text{Mn})\text{As}$  ferromagnetic semiconductors. *Appl. Phys. Lett.* **99**, 091111 (2011).
22. Matsubara, M. *et al.* Ultrafast photoinduced insulator-ferromagnet transition in the perovskite manganite  $\text{Gd}_{0.55}\text{Sr}_{0.45}\text{MnO}_3$ . *Phys. Rev. Lett.* **99**, 207401 (2007).
23. Ehrke, H. *et al.* Photoinduced melting of antiferromagnetic order in  $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$  measured using ultrafast resonant soft X-ray diffraction. *Phys. Rev. Lett.* **106**, 217401 (2011).
24. Miyasaka, K. *et al.* Ultrafast photoinduced magnetic moments in a charge-orbital-ordered antiferromagnetic  $\text{Nd}_{0.5}\text{Sr}_{0.5}\text{MnO}_3$  thin film. *Phys. Rev. B* **74**, 012401 (2006).
25. Wang, J. *et al.* Ultrafast enhancement of ferromagnetism via photoexcited holes in  $\text{GaMnAs}$ . *Phys. Rev. Lett.* **98**, 217401 (2007).
26. Kagan, M. Y., Khomskii, D. I. & Mostovoy, M. V. Double-exchange model: phase separation versus canted spins. *Eur. Phys. J. B* **12**, 217–223 (1999).
27. Cépas, O., Krishnamurthy, H. R. & Ramakrishnan, T. V. Instabilities and insulator-metal transitions in half-doped manganites induced by magnetic-field and doping. *Phys. Rev. B* **73**, 035218 (2006).
28. Yamaguchi, S., Okimoto, Y., Ishibashi, K. & Tokura, Y. Magneto-optical Kerr effects in perovskite-type transition-metal oxides:  $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$  and  $\text{La}_{1-x}\text{Sr}_x\text{CoO}_3$ . *Phys. Rev. B* **58**, 6862–6870 (1998).
29. Axt, V. M. & Mukamel, S. Nonlinear optics of semiconductor and molecular nanostructures; a common perspective. *Rev. Mod. Phys.* **70**, 145–174 (1998).
30. Karadimitriou, M. E., Kavousanaki, E. G., Dani, K. M., Fromer, N. A. & Perakis, I. E. Strong electronic correlation effects in coherent multidimensional nonlinear optical spectroscopy. *J. Phys. Chem. B* **115**, 5634–5647 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the National Science Foundation (contract no. DMR-1055352). Material synthesis at the Ames Laboratory was supported by the US Department of Energy-Basic Energy Sciences (contract no. DE-AC02-7CH11358). L.M. acknowledges FP7-REGPOT-2008-1, Project BIOSOLENTI no. 229927 for financial support. L.M. and I.E.P. were partially supported by the EU Social Fund and National resources through the THALES program NANOPHOS.

**Author Contributions** T.L., A.P. and J.W. performed the experimental measurements, collected the data and analysed the results. L.M. and I.E.P. developed the theory and performed the time-dependent and band-structure calculations. J.Y. and T.A.L. grew the samples and characterized the single crystals. J.W. and I.E.P. designed the experiment and wrote the paper, with help from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W. ([jgwang@iastate.edu](mailto:jgwang@iastate.edu)) and I.E.P. ([ilias@physics.uoc.gr](mailto:ilias@physics.uoc.gr)).

# Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution

Chien-Chun Chen<sup>1,2\*</sup>, Chun Zhu<sup>1,2\*</sup>, Edward R. White<sup>1,2</sup>, Chin-Yi Chiu<sup>2,3</sup>, M. C. Scott<sup>1,2</sup>, B. C. Regan<sup>1,2</sup>, Laurence D. Marks<sup>4</sup>, Yu Huang<sup>2,3</sup> & Jianwei Miao<sup>1,2</sup>

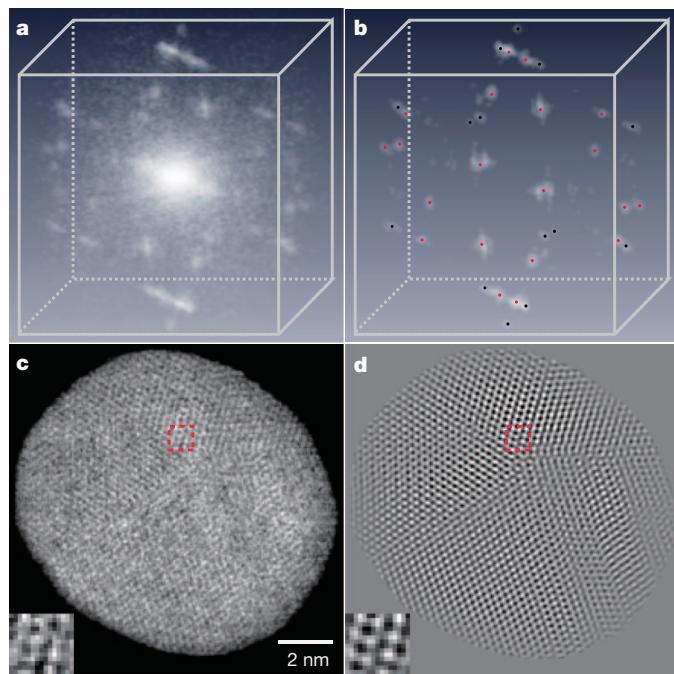
**Dislocations and their interactions strongly influence many material properties, ranging from the strength of metals and alloys to the efficiency of light-emitting diodes and laser diodes<sup>1–4</sup>. Several experimental methods can be used to visualize dislocations. Transmission electron microscopy (TEM) has long been used to image dislocations in materials<sup>5–9</sup>, and high-resolution electron microscopy can reveal dislocation core structures in high detail<sup>10</sup>, particularly in annular dark-field mode<sup>11</sup>. A TEM image, however, represents a two-dimensional projection of a three-dimensional (3D) object (although stereo TEM provides limited information about 3D dislocations<sup>4</sup>). X-ray topography can image dislocations in three dimensions, but with reduced resolution<sup>12</sup>. Using weak-beam dark-field TEM<sup>13</sup> and scanning TEM<sup>14</sup>, electron tomography has been used to image 3D dislocations at a resolution of about five nanometres (refs 15, 16). Atom probe tomography can offer higher-resolution 3D characterization of dislocations, but requires needle-shaped samples and can detect only about 60 per cent of the atoms in a sample<sup>17</sup>. Here we report 3D imaging of dislocations in materials at atomic resolution by electron tomography. By applying 3D Fourier filtering together with equal-slope tomographic reconstruction, we observe nearly all the atoms in a multiply twinned platinum nanoparticle. We observed atomic steps at 3D twin boundaries and imaged the 3D core structure of edge and screw dislocations at atomic resolution. These dislocations and the atomic steps at the twin boundaries, which appear to be stress-relief mechanisms, are not visible in conventional two-dimensional projections. The ability to image 3D disordered structures such as dislocations at atomic resolution is expected to find applications in materials science, nanoscience, solid-state physics and chemistry.**

Crystallographic techniques are currently the primary means for determining the 3D atomic structures of crystals. They determine atomic positions by averaging over many unit cells. As a complement to crystallographic techniques, electron tomography is an ideal technique for imaging the 3D local structure of materials at high resolution<sup>15,16,18–20</sup>. By combining annular dark-field scanning TEM with the centre-of-mass (CM) and equally sloped tomography (EST) methods, electron tomography has recently achieved a resolution of 2.4 Å (ref. 20). However, dynamical scattering effects<sup>21</sup>, the missing-wedge problem<sup>16,18,20</sup> and Poisson noise in the tilt series introduce noise in the EST reconstruction. Consequently, although lattice structure and some individual atoms are visible in the reconstruction<sup>20</sup>, electron tomography has not been able to reveal 3D dislocations in materials at atomic resolution. Here we overcome this obstacle by combining 3D Fourier filtering with high-angle annular dark-field scanning TEM (HAADF-STEM) tomography, and achieve 3D imaging of dislocations in a nanoparticle at atomic resolution.

Platinum (Pt) nanoparticles were synthesized by peptide sequences in aqueous solution (Methods)<sup>22</sup>. To make the nanoparticles more

stable under an electron beam, a thin (~1–2 nm) carbon layer was deposited on the nanoparticles (Methods). Using HAADF-STEM<sup>14</sup>, we acquired a tilt series of projections from a Pt nanoparticle (Methods). Supplementary Figs 1 and 2 show the tilt series of 104 projections with equal-slope increments and a tilt range of  $\pm 72.6^\circ$ . To monitor beam-induced changes to the Pt nanoparticle, three 0° projections were measured while acquiring the tilt series (Supplementary Fig. 3). The consistency of these projections indicates that the lattice structure of the nanoparticle was stable throughout the experiment.

After performing background subtraction and CM alignment (Methods), the tilt series was reconstructed by the EST method<sup>20,23–26</sup> (Methods). Figure 1a shows the 3D Fourier transform of the reconstruction and Fig. 1c shows a 2.6-Å-thick central slice in the *x*-*y* plane, where the *z* axis is along the beam direction. Figure 1b shows the 3D Fourier transform after applying a 3D Fourier filter, and Fig. 1d shows the corresponding central slice. Red dots in Fig. 1b indicate Bragg peaks, and black dots indicate non-Bragg peaks. The insets in Fig. 1c and d show enlarged regions of the atomic positions before and after applying the 3D Fourier filter.



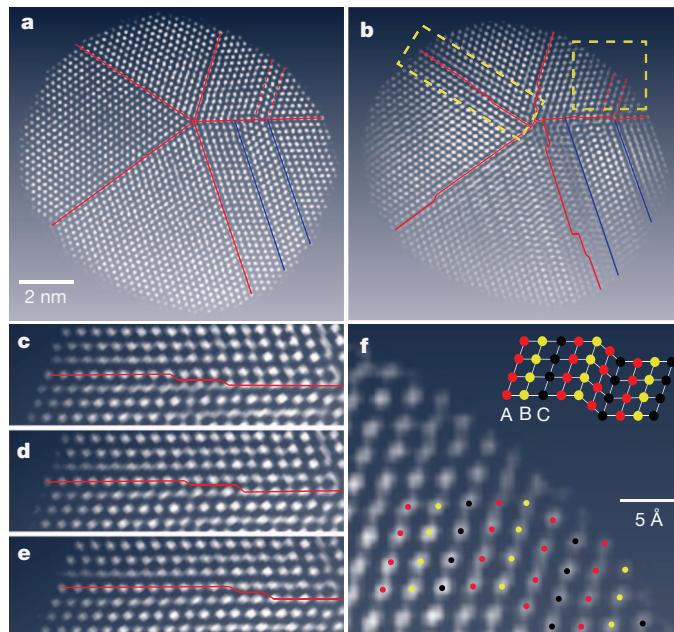
**Figure 1 | 3D reconstruction of a multiply twinned Pt nanoparticle before and after applying a 3D Fourier filter.** **a**, 3D Fourier transform of the raw reconstruction of the nanoparticle. **b**, 3D Fourier transform of the reconstruction after 3D Fourier filtering where the {111} and {200} Bragg peaks are labelled with red and black dots, respectively. **c**, A 2.6-Å-thick central slice in the *x*-*y* plane of the raw reconstruction, where the *z* axis is along the beam direction. **d**, The same slice of the 3D structure after applying a 3D Fourier filter, in which nearly all the atoms (in white) are visible. The clear boundary of the nanoparticle is due to the multiplication of the 3D structure with a 3D shape obtained from the EST reconstruction (Methods). The insets show an enlarged region of the atomic positions before and after applying a 3D Fourier filter.

<sup>1</sup>Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA. <sup>2</sup>California NanoSystems Institute, University of California, Los Angeles, California 90095, USA. <sup>3</sup>Department of Materials Science and Engineering, University of California, Los Angeles, California 90095, USA. <sup>4</sup>Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60201, USA.

\*These authors contributed equally to this work.

where the electron beam is along the  $z$  axis. However, owing to the low signal-to-noise ratio (SNR) in the EST reconstruction, 3D dislocations within the nanoparticle cannot be identified at atomic resolution. To enhance the SNR of the reconstruction, we developed a 3D Fourier filtering method to identify all the measurable 3D Bragg peaks and the 3D distribution around each peak (Methods). Figure 1b shows the 3D Fourier transform of the reconstruction after 3D Fourier filtering, in which the red and black dots indicate the {111} and {200} peaks of the Pt nanoparticle, respectively. By applying an inverse Fourier transform to Fig. 1b and multiplying it by the 3D shape of the Pt nanoparticle determined from the EST reconstruction, we obtained the 3D structure of the nanoparticle with a size of  $\sim 11.2 \times 10.7 \times 7.1 \text{ nm}^3$ . Supplementary Video 1 and Fig. 1d show 3D volume renderings and a 2.6-Å-thick central slice in the  $x$ - $y$  plane of the Pt nanoparticle, in which nearly all the atoms are visible.

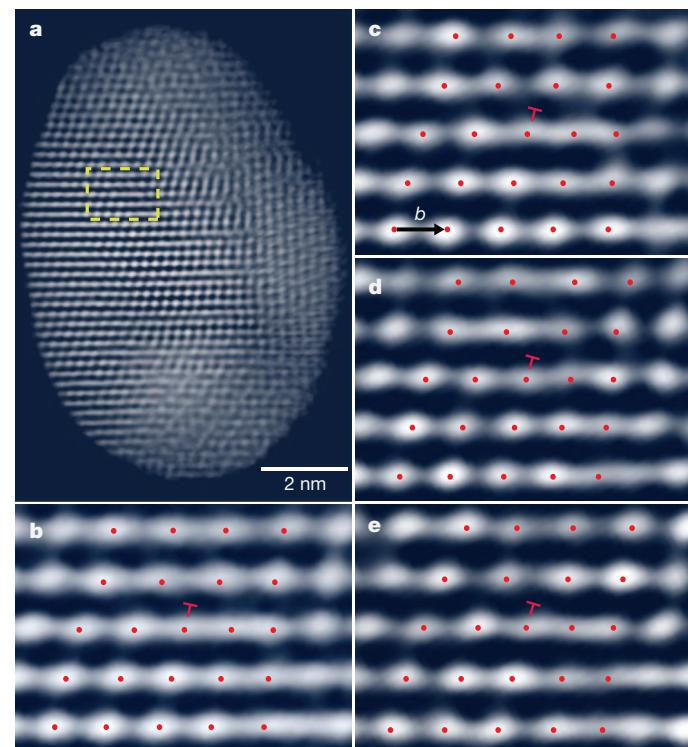
To verify the 3D Fourier filtering method, we performed a comparison with a 3D Wiener filter using the same experimental data. The Wiener filter is well established for reducing the noise in a signal, and it is applied to TEM images<sup>27</sup>. Supplementary Fig. 5b-d shows the 2.6-Å-thick central slice in the  $x$ - $y$  plane of the reconstruction after applying the 3D Wiener filter  $S^2/(S^2 + \lambda n^2)$ , where  $S$  is an estimate of the signal,  $n$  is the noise and  $\lambda$  is a parameter that controls the filtering strength (larger values of  $\lambda$  give stronger filtering). Supplementary Fig. 5f-j shows enlarged views of four regions for the raw reconstruction, and the reconstructions with the 3D Wiener filter ( $\lambda = 1, 2$  and  $3$ ) and the 3D Fourier filter. Although the result with the  $\lambda = 1$  Wiener filter is noisier, the atomic positions in the reconstructions using the  $\lambda = 2$  and  $3$  Wiener filters and the 3D Fourier filter are consistent.



**Figure 2 | Grain boundary comparisons between a 2D experimental projection and several 2.6-Å-thick internal slices of the reconstructed Pt nanoparticle.** **a**, Experimental projection in the  $x$ - $y$  plane suggesting that this is a decahedral multiply twinned nanoparticle and that the twin boundaries (red lines) are flat. Blue lines show two subgrain boundaries. To enhance the image contrast, a 2D Fourier filter was applied to the projection. **b**, A 2.6-Å-thick internal slice indicating the existence of atomic steps at the twin boundaries (red lines). The subgrain boundaries (blue lines) are two lattice spacings wider than those in **a**. **c**, Enlarged view of a twin boundary in **b**. **d** and **e**, a 2.6-Å-thick slice above and below the slice of **c**, revealing that the atomic steps vary in consecutive atomic layers. **f**, Enlarged view of a stacking fault in the 2.6-Å-thick internal slice, which is in good agreement with the classical model for a face-centred-cubic extrinsic stacking fault (inset). These images, as well as those in Figs 3 and 4, are displayed with Amira.

To further examine the 3D Fourier filtering method, we performed numerical simulations on a  $7.3 \times 7.0 \times 4.5 \text{ nm}^3$  decahedral Pt nanoparticle with multislice calculations. The Pt particle consists of 4,015 atoms with edge and screw dislocations (Supplementary Figs 6a-c, 7a and b). A tilt series of 63 projections with a tilt range of  $\pm 72.6^\circ$  and equal-slope increments was calculated by performing multislice STEM simulations (Supplementary Fig. 8). Two levels of Poisson noise were added to the projections of the tilt series with total electron doses of  $2.52 \times 10^5 \text{ e } \text{\AA}^{-2}$  and  $5.67 \times 10^4 \text{ e } \text{\AA}^{-2}$ , corresponding to  $R_{\text{noise}} = 10\%$  and  $20\%$ , respectively, where  $R_{\text{noise}}$  is an  $R$ -factor used to define the level of Poisson noise (Methods). The two tilt series were aligned and reconstructed by the CM and EST methods<sup>20,23-26</sup> (Methods, Supplementary Fig. 6d-f, j-l). Because of the low SNR in the reconstructions, not all the atoms or dislocations are visible in the raw 3D reconstructions, especially with  $R_{\text{noise}} = 20\%$  (Supplementary Fig. 6j-l). After applying a 3D Fourier filter to the raw reconstructions (Methods), we obtained two 3D structures of the simulated Pt nanoparticle with all the atoms resolved (Supplementary Fig. 6g-i, m-o). Furthermore, 3D grain boundaries, as well as the 3D core structures of edge and screw dislocations were determined at atomic resolution and are consistent with those in the model (Supplementary Figs 6g-i, m-o and 7c-f).

After verifying 3D Fourier filtering with a 3D Wiener filter and multislice simulation data, we analysed 3D dislocations of the Pt nanoparticle obtained from the experimental tilt series. Figure 2 shows grain boundary comparisons between a 2D experimental projection and 2.6-Å-thick internal slices of the reconstructed particle. The experimental projection in the  $x$ - $y$  plane suggests that this is a decahedral multiply



**Figure 3 | Observation of the 3D core structure of an edge dislocation at atomic resolution.** **a**, A 7.9-Å-thick internal slice of the nanoparticle. The lattice structure on the left and at the bottom parts of the slice is not well defined, mainly because this decahedral multiply twinned nanoparticle consists of five grains with different orientations. **b**, An enlarged view of an edge dislocation in **a** where red dots represent the position of the atoms. **c**, **d** and **e**, 2.6-Å-thick atomic layers sectioning through the slice of **b**. The three consecutive atomic layers indicate the dislocation line is in the direction of [101]. The Burgers vector ( $b$ ) of the edge dislocation was determined to be  $\frac{1}{2}[101]$ .

twinned nanoparticle<sup>28</sup> with flat twin boundaries (Fig. 2a and Supplementary Fig. 9). However, a 2.6-Å-thick internal slice in the  $x$ - $y$  plane and an enlarged view indicate the existence of atomic steps at the twin boundaries (Fig. 2b, c) that are hidden in the projection (Fig. 2a). Figure 2d and e shows enlarged views of a twin boundary in a 2.6-Å-thick slice above and below the slice of Fig. 2b, revealing that the atomic steps vary in consecutive atomic layers. These atomic steps are also independently verified by applying 3D Wiener filtering to the same experimental data (Supplementary Fig. 10). In addition, subgrain boundaries in the 2.6-Å-thick internal slice (Fig. 2b) are two lattice spacings wider than those in the projection (Fig. 2a). Figure 2f shows an enlarged view of a stacking fault in the 2.6-Å-thick internal slice ending at a twin boundary, which agrees well with the classical model for a face-centred-cubic extrinsic stacking fault<sup>1</sup> (inset in Fig. 2f).

In addition to twin boundaries, subgrain boundaries and stacking faults, we observed the 3D core structure of edge and screw dislocations at atomic resolution in the Pt nanoparticle. Figure 3a and b shows a 7.9-Å-thick internal slice of the nanoparticle and an enlarged view of an edge dislocation, where the red dots indicate the atomic positions. By computationally ‘sectioning’ the 7.9-Å-thick slice, we obtained three consecutive atomic layers, each 2.6-Å thick (Fig. 3c–e). The three consecutive atomic layers indicate that the dislocation line is in the [101] direction, and the Burgers vector of the edge dislocation was determined to be  $\frac{1}{2}[101]$  (Fig. 3c). To visualize a screw dislocation, a 5.3-Å-thick slice (two atomic layers) in the (111) plane was selected (Supplementary Fig. 11b) and then tilted to the [011] direction (Fig. 4a). Figure 4b shows an enlarged view of the slice where the zigzag pattern, a characteristic feature of a screw dislocation, is visible. To better visualize the screw dislocation, we display surface renderings of the enlarged region (Fig. 4c), where the atoms indicated by green dots are in the top layer and those indicated by red dots are in the bottom layer. The zigzag pattern is more clearly visualized in the surface renderings, in which the green line connects the atoms in the top layer and the red line connects the atoms in the bottom layer. The Burgers vector

of the screw dislocation was determined to be  $\frac{1}{2}[01\bar{1}]$ , and the width of the screw dislocation was estimated to be  $\sim 8.9$  Å, which is consistent with the results obtained by combining high-resolution TEM with image simulations for Au and Ir (ref. 29).

Careful analysis of the position of the screw dislocation inside the Pt nanoparticle suggests that the screw dislocation is associated with atomic steps at a twin boundary (Supplementary Fig. 11). Although it is often thought that nanoparticles cannot support dislocations, this is not the case for multiply twinned particles such as the decahedral one imaged here. These contain a  $\sim 2\%$  angular strain and a disclination; at larger sizes this can in principle be relieved by dislocations<sup>28</sup>. A recent analysis implied that about one-third of this strain is accommodated at the twin boundaries<sup>30</sup>, but as a 2D projection method was used in this analysis, it did not provide further insight. Our results strongly suggest that the twin boundaries are not flat and that dislocations associated with atomic steps at the boundaries account for the strain relaxation (Supplementary Fig. 11).

The significance of the present work is twofold. First, 3D atomic-resolution imaging of dislocations allows us to observe new structures that are not visible in conventional 2D projections; this is expected to advance our fundamental understanding of dislocations in materials. Second, although discrete tomography through the use of a priori information has been applied to reconstruct 3D surface morphology of a small crystalline nanoparticle at atomic resolution<sup>31</sup>, EST-based electron tomography in combination with 3D Fourier filtering represents a general method for 3D atomic resolution imaging of the local structure in nanomaterials. Although nanoparticles are used in this study, this method could, in principle, be applied to 3D imaging of thin materials at atomic resolution; the sample thickness is limited only by dynamical electron scattering.

## METHODS SUMMARY

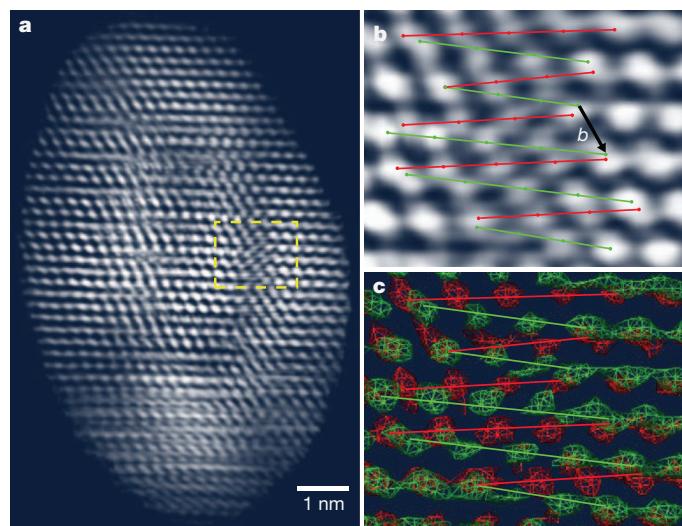
Pt nanoparticles were synthesized by using peptide sequences in aqueous solution at room temperature<sup>22</sup> and supported on 5-nm-thick silicon nitride membranes. To stabilize the nanoparticles under a STEM beam, a thin ( $\sim 1$ –2 nm) carbon layer was deposited on the Pt nanoparticles and the electron energy was kept below the knock-on radiation damage threshold of Pt. Using HAADF-STEM (energy: 200 keV; spherical aberration: 1.2 mm; illumination semi-angle: 10.7 mrad; detector inner and outer angles: 35.2 mrad and 212.3 mrad, respectively), a tilt series of 104 projections was acquired from a Pt nanoparticle with equal-slope increments<sup>20</sup> and a tilt range of  $\pm 72.6^\circ$ . To monitor beam-induced changes to the nanoparticle, three  $0^\circ$  projections were measured during the acquisition of the tilt series (Supplementary Fig. 3); they indicated that the Pt nanoparticle was stable throughout the experiment. After performing background subtraction and CM alignment for the tilt series, a 3D reconstruction of the nanoparticle was obtained using the EST method<sup>20,23–26</sup>. However, 3D dislocations of the nanoparticle cannot be identified in the raw 3D reconstruction at atomic resolution. To enhance the SNR of the reconstruction, we developed a 3D Fourier filtering method to identify all the measurable 3D Bragg peaks and the 3D distribution around each peak. The 3D Fourier transform of the EST reconstruction consists of {111} and {200} Bragg peaks. We first determined a two-shell volume with all the measurable Bragg peaks and set other voxels to zero. We then chose the most intense {111} Bragg peak as a reference peak. After optimizing the threshold based on the reference peak (Supplementary Fig. 4) and performing convolution with a three-voxel-diameter sphere, we identified all the measurable peaks and the 3D distribution around each peak (Fig. 1b). After applying the inverse Fourier transform to Fig. 1b, the 3D structure of the Pt nanoparticle was obtained in which nearly all the atoms are visible.

**Full Methods** and any associated references are available in the online version of the paper.

Received 26 September 2012; accepted 11 February 2013.

Published online 27 March 2013.

- Hull, D. & Bacon, D. J. *Introduction to Dislocations* 5th edn (Butterworth-Heinemann, 2011).
- Smith, W. F. & Hashemi, J. *Foundations of Materials Science and Engineering* 4th edn (McGraw-Hill Science, 2005).



**Figure 4 | Observation of the 3D core structure of a screw dislocation at atomic resolution.** **a**, Volume renderings of a 5.3-Å-thick slice (two atomic layers) in the (111) plane (Supplementary Fig. 11b), tilted to the [011] direction to visualize the zigzag pattern, a characteristic feature of a screw dislocation. **b**, Enlarged view of a screw dislocation showing the zigzag pattern. **c**, Surface renderings of the screw dislocation where the atoms represented by green dots are in the top layer and those by red dots are in the bottom layer. The zigzag pattern is more clearly visualized, the Burgers vector ( $b$ ) of the screw dislocation was determined to be  $\frac{1}{2}[01\bar{1}]$ , and the width of the screw dislocation was estimated to be  $\sim 8.9$  Å.

3. Nakamura, S. The roles of structural imperfections in InGaN-based blue light-emitting diodes and laser diodes. *Science* **281**, 956–961 (1998).
4. Hua, G. C. *et al.* Microstructure study of a degraded pseudomorphic separate confinement heterostructure blue-green laser diode. *Appl. Phys. Lett.* **65**, 1331–1333 (1994).
5. Hirsch, P. B., Horne, R. W. & Whelan, M. J. LXVIII. Direct observations of the arrangement and motion of dislocations in aluminium. *Phil. Mag.* **1**, 677–684 (1956).
6. Bollmann, W. Interference effects in the electron microscopy of thin crystal foils. *Phys. Rev.* **103**, 1588–1589 (1956).
7. Menter, J. W. The direct study by electron microscopy of crystal lattices and their imperfections. *Proc. R. Soc. Lond. A* **236**, 119–135 (1956).
8. Howie, A. & Whelan, M. J. Diffraction contrast of electron microscope images of crystal lattice defects. III. Results and experimental confirmation of the dynamical theory of dislocation image contrast. *Proc. R. Soc. Lond. A* **267**, 206–230 (1962).
9. Hirsch, P. B., Cockayne, D. J. H., Spence, J. C. H. & Whelan, M. J. 50 years of TEM of dislocations: past, present and future. *Phil. Mag.* **86**, 4519–4528 (2006).
10. Spence, J. C. H. *Experimental High-Resolution Electron Microscopy* 3rd edn (Oxford Univ. Press, 2003).
11. Chisholm, M. F. & Pennycook, S. J. Structural origin of reduced critical currents at  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$  grain boundaries. *Nature* **351**, 47–49 (1991).
12. Ludwig, W. *et al.* Three-dimensional imaging of crystal defects by ‘topotomography’. *J. Appl. Crystallogr.* **34**, 602–607 (2001).
13. Cockayne, D. J. H., Ray, I. L. F. & Whelan, M. J. Investigations of dislocation strain fields using weak beams. *Phil. Mag.* **20**, 1265–1270 (1969).
14. Pennycook, S. J. & Nellist, P. D. *Scanning Transmission Electron Microscopy: Imaging and Analysis* 1st edn (Springer, 2011).
15. Barnard, J. S., Sharp, J., Tong, J. R. & Midgley, P. A. High-resolution three-dimensional imaging of dislocations. *Science* **313**, 319 (2006).
16. Midgley, P. A. & Weyland, M. in *Scanning Transmission Electron Microscopy: Imaging and Analysis*. (eds Pennycook, S. J. & Nellist, P. D.) 353–392 (Springer, 2011).
17. Kelly, T. F. & Miller, M. K. Atom probe tomography. *Rev. Sci. Instrum.* **78**, 031101 (2007).
18. Xin, H. L., Ercius, P., Hughes, K. J., Engstrom, J. R. & Muller, D. A. Three-dimensional imaging of pore structures inside low- $\kappa$  dielectrics. *Appl. Phys. Lett.* **96**, 223108 (2010).
19. Bar Sadan, M. *et al.* Toward atomic-scale bright-field electron tomography for the study of fullerene-like nanostructures. *Nano Lett.* **8**, 891–896 (2008).
20. Scott, M. C. *et al.* Electron tomography at 2.4 Å resolution. *Nature* **483**, 444–447 (2012).
21. Howie, A. Diffraction channelling of fast electrons and positrons in crystals. *Phil. Mag.* **14**, 223–237 (1966).
22. Chiu, C. Y. *et al.* Platinum nanocrystals selectively shaped using facet-specific peptide sequences. *Nature Chem.* **3**, 393–399 (2011).
23. Miao, J., Föster, F. & Levi, O. Equally sloped tomography with oversampling reconstruction. *Phys. Rev. B* **72**, 052103 (2005).
24. Lee, E. *et al.* Radiation dose reduction and image enhancement in biological imaging through equally sloped tomography. *J. Struct. Biol.* **164**, 221–227 (2008).
25. Fahimian, B. P., Mao, Y., Cloetens, P. & Miao, J. Low dose X-ray phase-contrast and absorption CT using equally-sloped tomography. *Phys. Med. Biol.* **55**, 5383–5400 (2010).
26. Zhao, Y. *et al.* High resolution, low dose phase contrast x-ray tomography for 3D diagnosis of human breast cancers. *Proc. Natl Acad. Sci. USA* **109**, 18290–18294 (2012).
27. Marks, L. D. Wiener-filter enhancement of noisy HREM images. *Ultramicroscopy* **62**, 43–52 (1996).
28. Howie, A. & Marks, L. D. Elastic strains and the energy balance for multiply twinned particles. *Phil. Mag. A* **49**, 95–109 (1984).
29. Balk, T. J. & Hemker, K. J. High resolution transmission electron microscopy of dislocation core dissociations in gold and iridium. *Phil. Mag. A* **81**, 1507–1531 (2001).
30. Johnson, C. L. J. *et al.* Effects of elastic anisotropy on strain distributions in decahedral gold nanoparticles. *Nature Mater.* **7**, 120–124 (2008).
31. Van Aert, S., Batenburg, K. J., Rossell, M. D., Erni, R. & Van Tendeloo, G. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature* **470**, 374–377 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank B. S. Dunn for commenting on our manuscript and L. Ruan for discussions. The tomographic tilt series were acquired at the Electron Imaging Center for NanoMachines of California NanoSystems Institute. This work was supported by UC Discovery/TomoSoft Technologies (IT107-10166). L.D.M. acknowledges support by the NSF MRSEC (DMR-1121262) at the Materials Research Center of Northwestern University.

**Author Contributions** J.M. conceived and directed the project; C.-Y.C., C.Z., Y.H. and M.C.S. synthesized and prepared the samples; C.Z., E.R.W., B.C.R. and J.M. designed and conducted the experiments; C.-C.C. and J.M. performed the CM alignment and EST reconstruction; J.M., C.-C.C., C.Z. and L.D.M. analysed and interpreted the results; J.M., C.-C.C. and C.Z. wrote the manuscript. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Our EST reconstruction software is freely available at [www.physics.ucla.edu/research/imaging/EST](http://www.physics.ucla.edu/research/imaging/EST). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. (miao@physics.ucla.edu).

## METHODS

**Definition of  $R_{\text{noise}}$ .** An  $R$ -factor was used to define the level of Poisson noise in each multislice simulation tilt series,

$$R_{\text{noise}}^{\theta} = \frac{\sum_{x,y} |P_{\text{noise}}^{\theta}(x,y) - P_{\text{simulated}}^{\theta}(x,y)|}{\sum_{x,y} P_{\text{simulated}}^{\theta}(x,y)} \quad (1)$$

where  $P_{\text{simulated}}^{\theta}(x,y)$  is the projection calculated from multislice STEM simulations at angle  $\theta$ , and  $P_{\text{noise}}^{\theta}(x,y)$  is the same projection with Poisson noise added. After computing  $R_{\text{noise}}^{\theta}$  for each projection, we calculated  $R_{\text{noise}}$  by averaging  $R_{\text{noise}}^{\theta}$  for all the projections.

**Synthesis of Pt nanoparticles.** The Pt nanoparticles were synthesized by peptides in aqueous solution at room temperature, as reported previously<sup>22</sup>. All reagents were dissolved in water before using. A pre-prepared vial containing precursor (chloroplatinic acid hydrate ( $\text{H}_2\text{Pt}(\text{IV})\text{Cl}_6 \cdot x\text{H}_2\text{O}$ , 1 mM) and S7 peptide (Ser-Ser-Phe-Pro-Gln-Pro-Asn) solution (30 mg ml<sup>-1</sup>) were mixed with ascorbic acid (2 mM) immediately before injecting fresh  $\text{NaBH}_4$  (0.8 mM), where  $\text{NaBH}_4$  and ascorbic acid were used as reducing agents. The final volume of the reaction solution was 5 ml, and the reaction normally required more than 30 min.

**Sample preparation.** Pt nanoparticles were deposited on a 5-nm-thick silicon nitride membrane. The membrane, with a size of 100  $\mu\text{m} \times 1,500 \mu\text{m}$ , is supported on a 100- $\mu\text{m}$ -thick silicon frame (<http://TEMwindows.com>). To dissipate charge efficiently and make the nanoparticles more stable under an electron beam, a high-temperature ultrathin carbon coating (<http://TEMwindows.com>) was applied to the nanoparticles based on the following procedure. The silicon nitride membrane grid was first placed in a vacuum chamber and the temperature was increased from 300 °C to 700 °C at a rate of 10 °C s<sup>-1</sup>. The carbon was coated during a 5-min soak at 700 °C. The chamber then naturally cooled to 450 °C over the next 5 min before removing the grid. The silicon nitride membrane grid was finally loaded on a tomographic sample holder (Fischione 2020) for data acquisition.

**Acquisition of tomographic tilt series using HAADF-STEM.** STEM images of the Pt nanoparticles were acquired on a FEI Titan 80-300 microscope (energy: 200 keV; spherical aberration: 1.2 mm; illumination semi-angle, 10.7 mrad). The 100-pA electron beam was focused to a probe with a 50- $\mu\text{m}$  probe-forming aperture (C2) and rastered over the sample. The scattered electrons were captured by a Fischione 3000 HAADF detector with angles between 35.2 mrad and 212.3 mrad relative to the optical axis. HAADF angles were used to reduce the nonlinear intensities and the diffraction contrast in the images. The maximum tilt angles were limited by the holder to  $\pm 75^\circ$ . To reduce vibration and drift during data acquisition, the sample holder was allowed to settle for 1 h after inserting into the microscope and also for several minutes after moving to each new angle. Tilt series were manually acquired by changing the angle with equal-slope increments<sup>20,23–26</sup>. When focusing an image, a nearby nanoparticle (generally within 20 nm along the rotation axis) was first viewed, thus reducing unnecessary radiation dose to the particle under study<sup>20</sup>. Using this low-exposure acquisition scheme, a tomographic tilt series of 104 projections with equal-slope increments and a tilt range of  $\pm 72.6^\circ$  was acquired from a Pt nanoparticle. The probe current was  $\sim 100$  pA with a dwell time of 48  $\mu\text{s}$  per pixel, and the magnification of each projection was  $3.6 \times 10^6$ . The total electron dose of the tilt series was estimated to be  $\sim 2.5 \times 10^7$  e Å<sup>-2</sup>. Because the pixel size in STEM mode may vary, a calibration image of an oriented single-crystal Au foil (Ted Pella) was taken in STEM mode under the same conditions, and the STEM pixel size was calibrated to be 0.35 Å. To enhance the SNR in the projections of the tilt series, 1.5  $\times$  1.5 pixel binning was performed for each projection. The pixel size of the binned projections is 0.53 Å.

**Background subtraction and CM alignment.** To use the EST iterative algorithm, background subtraction has to be performed for each projection. We implemented background subtraction and CM alignment based on the following procedure. First, we selected the 0° projection and chose a suitable cut-off value to subtract its background. After projecting the 0° projection onto the  $y$ -axis (that is, the tilt axis) to obtain a one-dimensional (1D) curve, we calculated the centre of mass of the 1D curve ( $y_{\text{CM}}$ ), and set  $y_{\text{CM}}$  as the origin of the  $y$ -axis. This 1D curve was used as a reference curve. Second, we optimized the background subtraction for other projections based on the reference curve. We scanned the cut-off values from 0.5 to 1.5 times the mean value of each projection with an increment of 0.01. For each cut-off value, we projected the projection onto the  $y$ -axis to obtain a 1D curve. By shifting the 1D curve pixel-by-pixel along the  $y$ -axis, we calculated the difference between the 1D curve and the reference curve. We recorded the smallest difference and the corresponding shift for each 1D curve. After scanning through all the cut-off values, we plotted the smallest differences against the cut-off values, which should give a U-shaped curve. We identified the minimum corresponding to the optimized cut-off value and the shift of a given projection. After performing

background subtraction with the optimized cut-off values, we calculated the centre of mass ( $y_{\text{CM}}$ ) for all the 1D curves. If all of them (that is, rounded  $y_{\text{CM}}$ ) were not at the origin, we adjusted the cut-off value for the 0° projection (that is, the reference curve) and repeated the above steps until the best agreement was achieved. Finally, after performing background subtraction and aligning the projections along the  $y$ -axis, we projected all the projections onto the  $x$  axis. We calculated the centre of mass ( $x_{\text{CM}}$ ) for all 1D curves, and set  $x_{\text{CM}}$  as the origin of the  $x$ -axis<sup>20</sup>. After repeating this process, all the projections were aligned to the tilt axis.

**EST reconstruction.** The EST iterative algorithm started with converting the measured projections to Fourier slices by the fractional Fourier transform<sup>32</sup>. The algorithm then iterated alternately between real and reciprocal space through the use of the pseudopolar fast Fourier transform<sup>33</sup>. In real space, the voxel values outside a loose support (that is, a rectangular box larger than the true boundary of the structure to be reconstructed) and the negative voxel values inside the support were set to zero, whereas in reciprocal space the corresponding calculated Fourier slices were replaced with the measured ones and the remaining slices kept unchanged in each iteration. Each iteration was monitored by an error metric, defined as the difference between the measured and calculated Fourier slices, and the algorithm was terminated after reaching a maximum number of iterations. Using the iterative EST algorithm, a preliminary 3D reconstruction was obtained after 500 iterations. An updated 3D support was determined by convolving the reconstruction with a Gaussian window and selecting a suitable cut-off. The 3D shape of the support was also double checked by examining the reconstruction slice-by-slice to ensure the support does not crop the structure. Using the updated support, we performed another 500 iterations to obtain a new reconstruction. To further improve the 3D reconstruction, we also projected the reconstruction back to calculate projections at given angles. By computing the cross-correlation between the calculated and measured projections, we further adjusted the alignment of the projections to achieve maximum consistency in 3D reconstruction. Usually the shift should be one pixel or smaller in each dimension. Otherwise, the data analysis and CM alignment procedure has to be re-done. We then repeated the procedure for improving the support and back-projection alignment. The final reconstruction was obtained when no further improvements can be made. The details of the EST algorithm can be found in refs 20, 23–26.

**3D Fourier filtering.** As a result of dynamical scattering effects<sup>21</sup>, the missing-wedge problem<sup>16,18,20</sup> and Poisson noise in the tilt series, not all the atoms or dislocations are visible in the raw 3D EST reconstruction. To enhance the SNR in the reconstruction, we developed a 3D Fourier filtering method by using the following procedure. First, the 3D Fourier transform of the raw reconstruction of the Pt nanoparticle consists of two sets of lattice planes {111} and {200}. The intensities of the {111} peaks were estimated to be several times higher than those of the {200} peaks. We calculated the average radial distance ( $d$ ) between the {111} and {200} peaks. Two radii were then determined by  $R_{\text{in}} = R_{111} - d$  and  $R_{\text{out}} = R_{200} + d$ , where  $R_{111}$  and  $R_{200}$  are the average radial distance for the {111} and {200} peaks, respectively. By keeping those voxels in the 3D Fourier transform with their radii between  $R_{\text{in}}$  and  $R_{\text{out}}$ , and setting other voxels to zero, we obtained a two-shell volume including all the measurable 3D Bragg peaks.

Next, we implemented a method to further reduce noise among the Bragg peaks within the two-shell volume. We chose the most intense {111} Bragg peak as a reference peak and calculated thresholds based on the reference peak. We scanned the thresholds from 1% to 20% of the reference peak in steps of 1%. For each threshold, we set voxels with values larger than the threshold to one and other voxels to zero, and obtained a 3D mask. The 3D mask was convolved with a three-voxel-diameter sphere to compute a new 3D mask, where the convolution process was to retain the 3D distribution of each Bragg peak. By multiplying the new 3D mask with the Fourier transform of the raw reconstruction, we obtained a new 3D Fourier transform. By monitoring the change in the noise among the Bragg peaks, we found that a threshold with 10% of the reference peak is large enough to remove noise among the 3D Bragg peaks, while retaining all the measurable {111} and {200} peaks and the 3D distribution around each peak (Fig. 1b and Supplementary Fig. 4). The optimized threshold of 10% of the reference peak obtained here may vary for different samples.

Finally, by applying the inverse Fourier transform to Fig. 1b and multiplying it by a 3D shape (that is, a tight support) obtained from the EST reconstruction, we obtained the 3D structure of the Pt nanoparticle (Supplementary Video 1 and Fig. 1d). We confirmed the accuracy of the 3D Fourier filtering method by using two independent approaches: (1) multislice STEM simulations<sup>34</sup> of a decahedral Pt nanoparticle with edge and screw dislocations (Supplementary Figs 6–8) and (2) performing a comparison with a 3D Wiener filter<sup>35</sup> on the same experimental data (Supplementary Figs 5 and 10). In our numerical simulations, we also found that, compared to 2D Fourier filtering method<sup>36–38</sup>, 3D Fourier filtering is more accurate. This is because in 3D Fourier filtering, each voxel in 3D reciprocal space is correlated to all voxels in 3D real space, and vice versa. But in 2D Fourier filtering,

- the correlated information only exists in two dimensions. Thus, for a given object, there is more correlated information (voxels) in 3D Fourier filtering than in the 2D case.
32. Bailey, D. H. & Swarztrauber, P. N. The fractional Fourier transform and applications. *SIAM Rev.* **33**, 389–404 (1991).
  33. Averbuch, A., Coifman, R. R., Donoho, D. L., Israeli, M. & Shkolnisky, Y. A framework for discrete integral transformations I — the pseudopolar Fourier transform. *SIAM J. Sci. Comput.* **30**, 785–803 (2008).
  34. Kirkland, E. J. *Advanced Computing in Electron Microscopy* 2nd edn (Springer, 2010).
  35. Brown, R. G. & Hwang, P. Y. C. *Introduction to Random Signals and Applied Kalman Filtering* 3rd edn (Wiley, 1996).
  36. Saxton, W. O. *Computer Techniques for Image Processing in Electron Microscopy* (Academic, 1978).
  37. Hawkes, P. W. *Computer Processing of Electron Microscope Images* (Springer, 1980).
  38. Möbus, G., Necker, G. & Rühle, M. Adaptive Fourier-filtering technique for quantitative evaluation of high-resolution electron micrographs of interfaces. *Ultramicroscopy* **49**, 46–65 (1993).

# Colossal injection of catalyst atoms into silicon nanowires

Oussama Moutanabbir<sup>1</sup>, Dieter Isheim<sup>2</sup>, Horst Blumtritt<sup>3</sup>, Stephan Senz<sup>3</sup>, Eckhard Pippel<sup>3</sup> & David N. Seidman<sup>2</sup>

The incorporation of impurities during the growth of nanowires from the vapour phase alters their basic properties substantially, and this process is critical in an extended range of emerging nanometre-scale technologies<sup>1–4</sup>. In particular, achieving precise control of the behaviour of group III and group V dopants has been a crucial step in the development of silicon (Si) nanowire-based devices<sup>5–7</sup>. Recently<sup>8–11</sup> it has been demonstrated that the use of aluminium (Al) as a growth catalyst, instead of the usual gold, also yields an effective p-type doping, thereby enabling a novel and efficient route to functionalizing Si nanowires. Besides the technological implications, this self-doping implies the detachment of Al from the catalyst and its injection into the growing nanowire, involving atomic-scale processes that are crucial for the fundamental understanding of the catalytic assembly of nanowires. Here we present an atomic-level, quantitative study of this phenomenon of catalyst dissolution by three-dimensional atom-by-atom mapping of individual Al-catalysed Si nanowires using highly focused ultraviolet-laser-assisted atom-probe tomography. Although the observed incorporation of the catalyst atoms into nanowires exceeds by orders of magnitude the equilibrium solid solubility<sup>12</sup> and solid-solution concentrations in known non-equilibrium processes<sup>13,14</sup>, the Al impurities are found to be homogeneously distributed in the nanowire and do not form precipitates or clusters. As well as the anticipated effect on the electrical properties, this kinetics-driven colossal injection also has direct implications for nanowire morphology. We discuss the observed strong deviation from equilibrium using a model of solute trapping at step edges, and identify the key growth parameters behind this phenomenon on the basis of a kinetic model of step-flow growth of nanowires. The control of this phenomenon provides opportunities to create a new class of nanoscale devices by precisely tailoring the shape and composition of metal-catalysed nanowires.

The nanowires investigated in this study were grown on Si(111) in an ultrahigh-vacuum chemical deposition system<sup>10,15</sup>. Their growth was accomplished with Al-Si nanoparticles, which act catalytically as the energetically favoured sites for vapour-phase reactant ( $\text{SiH}_4$ ) adsorption and (when saturated) as the nucleation sites for crystallization and one-dimensional growth<sup>16</sup>. The resulting nanowires are aligned along the [111] direction and display morphological characteristics that are sensitive to the growth temperature (Supplementary Figs 1 and 2). Growth at the highest temperature (470 °C) yields tapered nanowires with an average height and base diameter of  $475 \pm 30$  nm and  $94 \pm 8$  nm, respectively (Supplementary Fig. 1). Figure 1a displays a high-resolution transmission electron microscopy (TEM) micrograph of an Al-catalysed Si nanowire near the interface with the catalyst. This micrograph demonstrates that both nanowire and catalyst nanoparticle are single crystals with a coherent interface between them. However, the close atomic numbers of  $^{13}\text{Al}$  and  $^{14}\text{Si}$  make them indistinguishable in high-resolution TEM images of nanowires. Energy-dispersive X-ray analysis (EDX) in an aberration-corrected TEM did not find a significant Al signal in Si nanowires

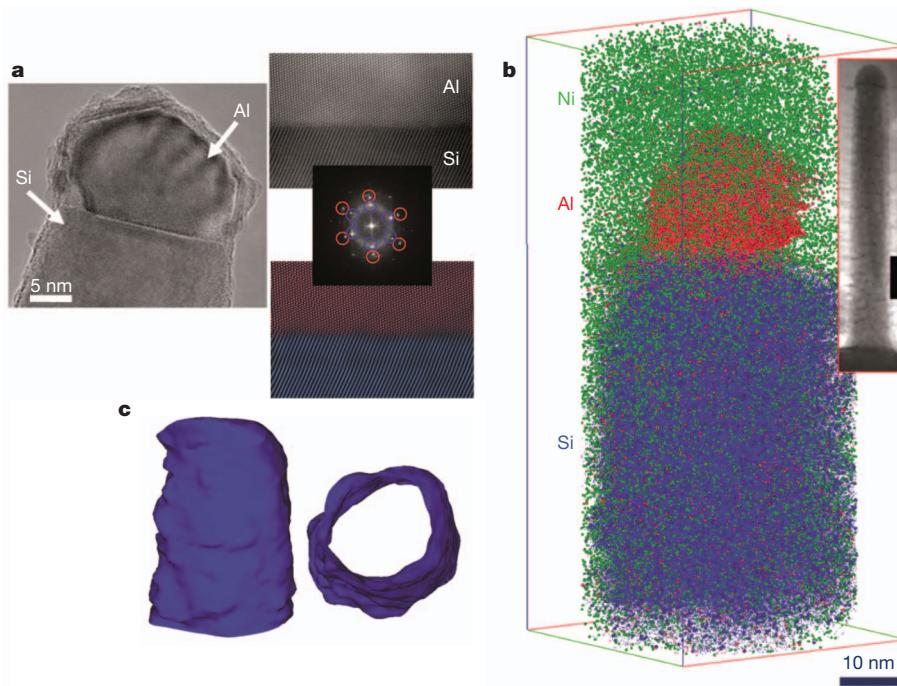
grown at the lowest temperature (410 °C). This indicates that the Al concentration is less than the EDX detection limit, which is typically  $\sim 0.5$  atomic per cent (at.%). For a rigorous analysis, we used atom-probe tomography (APT), which has a superior detection sensitivity of less than ten atomic parts per million (p.p.m.), with atomic-scale spatial resolution<sup>17</sup>. Pulsed evaporation of individual atoms was achieved using a focused picosecond ultraviolet (wavelength  $\lambda = 355$  nm) laser. The use of a highly focused ultraviolet laser beam decreases local heating, thereby improving the detection limits and mass resolving power ( $m/\Delta m$ , where  $m$  is the mass of an isotope)<sup>18</sup>.

Figure 1b shows a three-dimensional atom-by-atom map of a single Al-catalysed Si nanowire covered by a protective nickel (Ni) layer. The inset in Fig. 1b displays a cross-sectional TEM image of an identical nanowire. Figure 1c shows two projections of an 80-nm-long segment of a Si nanowire delineated by an isoconcentration surface drawn at the 50 at.% Si level.

The 50 at.% Al isoconcentration surface of the catalyst nanoparticle is displayed in Fig. 2a. Axial and radial concentration profiles of Al and Si, taken from the top 10 nm of the nanoparticle, are presented in Fig. 2b and c, respectively. Both concentration profiles demonstrate Si segregation close to the nanoparticle's surface, possibly as a result of Si expulsion during post-growth cooling or simply due to oxidation on exposure to air. The average Si concentration measured at the core of the nanoparticle is in the range of  $2.7 \pm 0.9$  to  $3.1 \pm 0.4$  at.%, which is nearly twice the solubility of Si in Al at the eutectic temperature<sup>19</sup>. This indicates that Si solubility in Al increases in undercooled catalyst. Figure 2d displays the three-dimensional distribution of Al atoms detected in an as-grown Si nanowire. The average Al solid concentration ( $x_{\text{Al}}^S$ ) measured for Si nanowires grown at 410 °C is  $\sim (2.0 \pm 0.5) \times 10^{20}$  atom cm<sup>-3</sup> ( $\sim 0.4$  at.%). Strikingly, this is about four orders of magnitude greater than the extrapolated equilibrium solubility of Al in Si (ref. 12), which corresponds to an increase of two to three orders of magnitude compared to Al-assisted recrystallization<sup>13</sup> or solid-phase epitaxy<sup>14</sup> at a similar temperature. Notwithstanding this colossal Al concentration, clusters or precipitates of Al do not form. To verify this quantitatively, we statistically analysed the spatial distribution of Al atoms (Fig. 2e) and performed partial Al-Al radial distribution function analyses. This analysis shows that the concentration frequency distribution is not different from a binomial distribution with the same mean at 99% confidence, and thus confirms that Al does not form precipitates or clusters in Si nanowires.

Figure 3a displays a set of radial concentration profiles measured in 8-nm-diameter cylinders in different regions across the nanowire grown at 410 °C, perpendicular to the growth direction (Fig. 3a insets). The estimated mean  $x_{\text{Al}}^S$  value varies between 0.39 and 0.45 at.%. The profiles display random fluctuations indicative of a random distribution of Al in Si nanowires. This is also consistent with the Al axial concentration profiles (Fig. 3b). Indeed,  $x_{\text{Al}}^S$  averaged over a cylindrical volume remains practically the same regardless of the diameter of the selected volume, thereby confirming the uniform distribution of

<sup>1</sup>Department of Engineering Physics, École Polytechnique de Montréal, Montréal, CP 6079, Succursale Centre-Ville, Montréal, Québec H3C 3A7, Canada. <sup>2</sup>Department of Materials Science and Engineering, and Northwestern University Center for Atom-Probe Tomography (NUCAPT), Northwestern University, Evanston, Illinois 60208-3108, USA. <sup>3</sup>Max Planck Institute of Microstructure Physics, Weinberg 2, 06120 Halle (Saale), Germany.



**Figure 1 | Structure and three-dimensional map of Al-catalysed Si nanowires.** **a**, High-resolution cross-sectional TEM image (left) displaying the interface between the catalyst particle and the nanowire. The right panel exhibits a close-up image of the interface (top), the fast Fourier transform of the image (middle), and the corresponding colour-filtered image (bottom, Al and Si regions correspond to red and blue regions, respectively) indicating that the interface is epitaxial. **b**, Three-dimensional APT atom-by-atom map of a

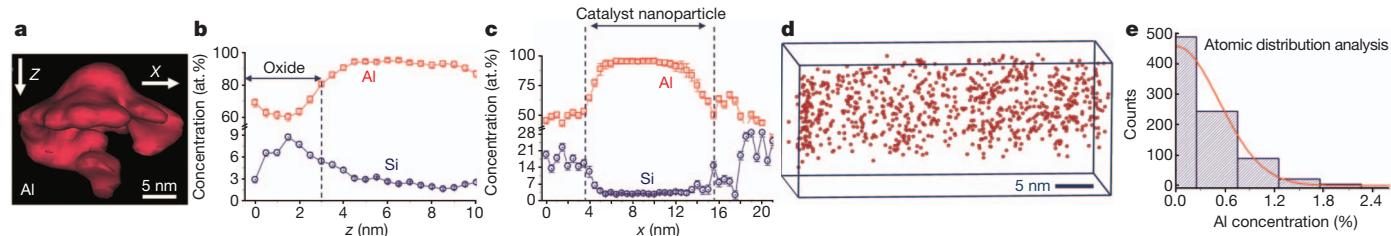
nanowire grown at 410 °C. For the sake of clarity, only a limited number of atoms is displayed ( $2.5 \times 10^4$  atoms of each element). Inset, a cross-sectional TEM image of an identical Si nanowire (scale bar, 40 nm). **c**, Si 50 at.% isoconcentration surface of an 80-nm-long segment of a nanowire determined by analysing a three-dimensional atom-probe tomographic reconstruction: left, side view; right, top view.

Al. Analysis of  $x_{\text{Al}}^{\text{S}}$  in Si nanowires grown at a higher temperature (470 °C) demonstrates a significantly stronger incorporation of Al, leading to concentrations sufficiently large to be detected by EDX (Fig. 3c). Interestingly, the Al radial concentration displays flat profiles, indicating a uniform distribution of Al in the nanowire. Surface segregation of Al would have led to ‘U’-shaped profiles. Similarly, the Al axial concentration profile indicates a uniform concentration (~4.3 at.%) along the nanowire. We note the ~10-fold increase in  $x_{\text{Al}}^{\text{S}}$  when compared to the nanowires grown at 410 °C.

It is generally recognized that nanowire tapering is induced by vapour–solid growth on sidewall surfaces<sup>8</sup> and/or catalyst diffusion from one nanowire to another<sup>20</sup>. As demonstrated here, this is not the case for Al-catalysed nanowires, for which the observed incorporation of catalyst atoms during growth is the key element determining nanowire morphology. It is noteworthy that the uniform Al radial concentration profile (Fig. 3c) rules out vapour–solid deposition on the nanowire sidewalls as a possible mechanism for nanowire tapering. Specifically, if homoepitaxy did take place on the nanowire’s sidewalls,

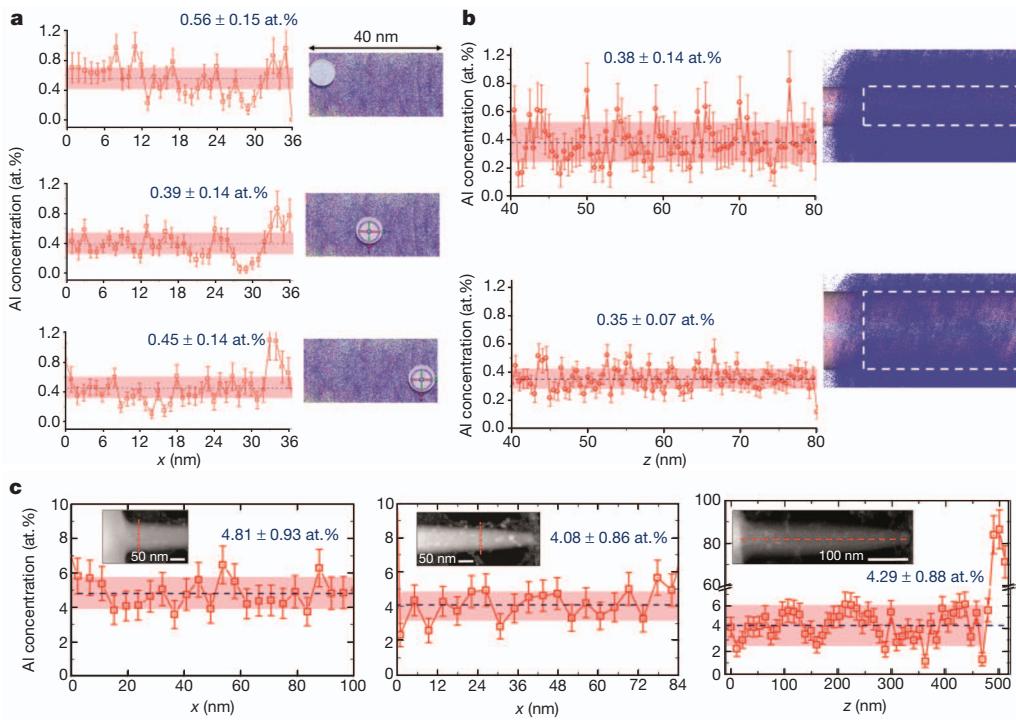
the overgrown layer would have been Al-free, because Al diffusion is negligible under these growth conditions<sup>21</sup>. The strong incorporation of Al catalyst atoms into the nanowire that we observe should involve a continuous dissolution of the catalyst throughout the growth process, with direct implications for nanowire morphology (Supplementary Information, section D). For instance, at 470 °C, the total volume of Al atoms incorporated in the completely grown Si nanowire is  $\sim 5.5 \times 10^4 \text{ nm}^3$ , which is significantly larger than the volume of the catalyst at the end of growth ( $\sim 1.3 \times 10^4 \text{ nm}^3$ ). This means that there is more Al in the Si nanowire than in the nanoparticle at the end of the growth process—that is, the catalyst has lost more than 80% of its initial volume. This leads to a continuous reduction of the diameter of the catalyst nanoparticle during growth, and thus to nanowire tapering. We note, however, that the catalyst’s volume loss is very small (~6%) at 410 °C, which explains the absence of tapering at this temperature (Supplementary Fig. 2).

The observed uniform Al concentration distributions suggest that the dissolution of Al atoms and their injection into the growing nanowire



**Figure 2 | APT data obtained for a catalyst nanoparticle and a Si nanowire.** **a**, Al 50 at.% isoconcentration surface of a catalyst nanoparticle. **b, c**, Axial (b) and radial (c) concentration profiles of Al and Si in the catalyst nanoparticle; error bars,  $\pm 1$  s.d. calculated for the analysed volume containing  $2.9 \times 10^5$  Al

atoms and  $8 \times 10^3$  Si atoms. **d**, Three-dimensional distribution of Al atoms in a Si nanowire grown at 410 °C. **e**, Results of statistical analysis of Al concentration frequency (histogram) and of binomial analysis (solid line). The binomial analysis was performed by sampling blocks containing 200 atoms.



**Figure 3 | Al concentration profiles in individual nanowires.** **a**, Al radial concentration profiles (left) measured in different regions of a Si nanowire. Each concentration profile is taken from a cylindrical segment having a diameter of 8 nm, as indicated in the images (right). **b**, Al axial concentration profiles (left) measured for cylindrical segments (right) with a diameter of 10 nm (top) and 20 nm (bottom). The horizontal dashed line in all concentration profiles denotes the average concentration, which is given as

blue numerals; the shaded band is the uncertainty interval; and error bars in **a** and **b** represent  $\pm 1\text{s.d.}$ . The number of Al atoms in the sampling volumes varies between  $\sim 2,000$  and  $4,500$ . **c**, EDX Al concentration profiles: left and middle plots, radial concentration profiles near the base and middle of the Si nanowire, respectively; right plot, axial concentration profile. Error bars, approximately  $\pm 0.5\text{ at.\%}$ , which is the sensitivity of the instrument. Insets, the corresponding TEM images with EDX scan lines (dashed lines).

occurs at a constant rate during the growth process. Moreover, the fact that Al displays uniform radial and axial concentration profiles suggests the existence of a single atomic pathway behind this colossal incorporation. We anticipate that surface effects, capillarity and related nanoscale stresses will shift the boundaries in the phase diagram of the catalyst–nanowire system in comparison to those of a bulk system<sup>22</sup>. However, by taking into account these thermodynamic considerations and including the nanowire tapering (Supplementary Information, section E), we found that the calculated solid composition of the nanowire is always smaller than the equilibrium content, which is of the order of 1 atomic p.p.m. at the growth temperature of our Si nanowires (Supplementary Fig. 11). This stands in sharp contrast to the observed excess concentrations (that is, concentrations above the equilibrium value) of orders of magnitude in our Si nanowires, which demonstrates that kinetic factors must be responsible for the incorporation of Al atoms.

Supersaturations of several orders of magnitude occur as a result of deviations from local equilibrium in ultra-rapid solidification of a melt<sup>23–25</sup>. This phenomenon (so-called solute trapping) has been observed for a variety of solutes in Si for growth rates of the order of metres per second (refs 24, 25). In general, solute trapping is quantified in terms of a partition coefficient,  $k$ , which is the ratio of the solute concentration in the solvent to its concentration in the melt at the solid–liquid interface, as a function of the equilibrium partition coefficient ( $k_e$ ), the average interface velocity ( $v_{\min}$ ), and a characteristic velocity known as diffusive velocity ( $v_D$ )<sup>25</sup>. Because nanowires grow by successive addition of bilayers through step flow<sup>26</sup>, the average interface velocity is defined as  $v_{\min} = b/\tau$ , where  $b = 0.3\text{ nm}$  is the thickness of a Si(111) bilayer and  $\tau$  is the total time needed for one bilayer to grow ( $\tau$  includes the time needed to reach the supersaturation (the incubation time,  $\tau_1$ ) and the time between the end of the incubation period (that is, the beginning of nucleation) and the complete growth of the bilayer,  $\tau_{BL}$ ). Note that the incubation time is a nanoscale phenomenon, not

expected for macroscopic systems<sup>26</sup>. Under our growth conditions,  $v_{\min}$  equals  $0.4$  and  $0.2\text{ nm s}^{-1}$  at  $470$  and  $410\text{ }^\circ\text{C}$ , respectively. Intriguingly, from studies of rapid solidification of melts<sup>25</sup>, one can infer that solute trapping should not occur for systems with an interface velocity of the order of nanometres per second, and thus the anticipated amount of Al in Si nanowires should not exceed the equilibrium concentration. We note that the influence of temperature on the rate of the thermally activated jumps is not sufficient to explain the observed incorporation of Al at exceedingly low interface velocities (Supplementary Information, section F).

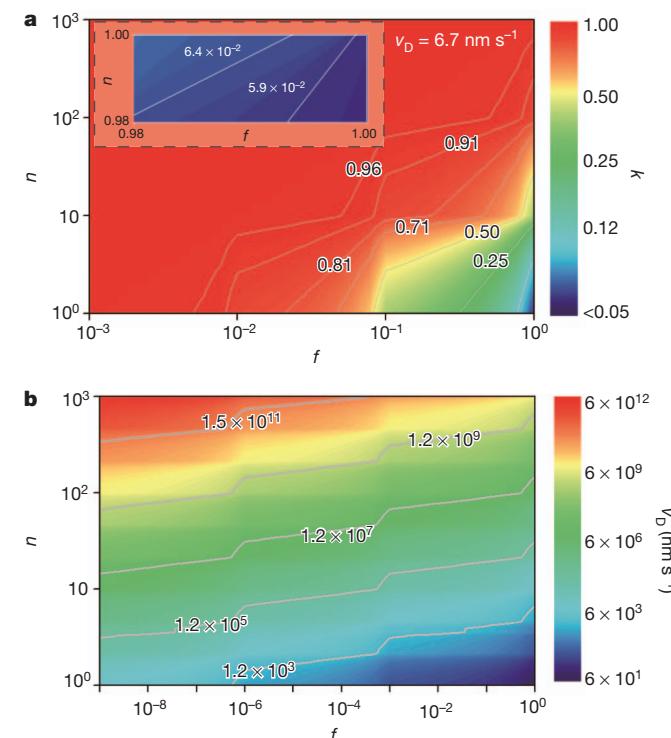
Current solute trapping models<sup>25</sup> involve the assumption that atomic jumps occur over the entire infinite liquid–solid interface, and solute atoms are trapped only after the passage of the next layer across the nanowire surface. Although this mechanistic picture describes the behaviour in bulk systems very well<sup>25</sup>, it cannot be extended to nanowire growth, which is characterized by a relatively large delay between two successive bilayers. During this incubation time, there is no driving force for atoms to leave the catalyst and thus the exchange (that is, the process of atomic jumps) should stop until the system reaches supersaturation. It is also important to note that in step-flow growth, atoms at the step edges have few nearest neighbours, which provides them with more steric freedom for an exchange process. Thus, it is reasonable to assume that atomic jumps are more favourable at the step edges than at terraces. This suggests that the impurity atoms become frozen in the solid immediately after the formation of the next row of atoms at the step edge. Therefore, the time,  $\tau_e$ , during which the local exchange is possible corresponds to the average time needed to add one row, and can be expressed as  $\tau_e = a/v_{step}$ , where  $a$  is equivalent to the width of one row of atoms and  $v_{step}$  is the average step velocity. This latter can be expressed as a function of nanowire radius  $R_{NW}$ :  $v_{step} = 2R_{NW}/\tau_{BL}$ . Thus, at fixed radius,  $\tau_e = \tau_{BL}(a/2R_{NW}) = \tau_{BL}/n$ , where  $n = 2R_{NW}/a \gg 1$ . The parameter  $n$  depends

only on step speed (Supplementary Information, section F). Experimentally, we can only determine the average time between the completion of one bilayer and the completion of the next,  $\tau = \tau_1 + \tau_{BL}$ . By taking  $f = \tau_{BL}/\tau$ , the exchange time can be written as  $\tau_e = (f/n)\tau$ .

The transport of impurity atoms between the two sides of the interface for  $0 < t < \tau_e$  can be described on the basis of chemical rate theory and mass conservation<sup>27,28</sup>. Assuming that the step growth is random rather than periodic<sup>29</sup> (Supplementary Information section F), the partition coefficient can be expressed as:

$$k = \frac{k_e + \frac{b^2}{D_i \tau} \times \frac{n}{f}}{1 + \frac{b^2}{D_i \tau} \times \frac{n}{f}} = \frac{k_e + \frac{v_{min}}{v_D} \times \frac{n}{f}}{1 + \frac{v_{min}}{v_D} \times \frac{n}{f}}$$

where  $D_i$  is the coefficient of interdiffusion at the interface ( $v_D = D_i/b$ ). Figure 4a displays the evolution of the calculated  $k$  in  $(f;n)$  space for nanowires grown at 470 °C ( $\tau = 0.75$  s and  $k_e = 2.6 \times 10^{-6}$ ) using  $v_D = 6.7 \text{ nm s}^{-1}$  obtained from earlier solute trapping models (Supplementary Information section F). We note that any pair  $(f;n)$  (except around  $(1;1)$ ) predicts Al trapping at levels always significantly higher than the measured value ( $k = 5.6 \times 10^{-2}$ ). At these trapping levels, nanowire growth cannot occur. Obviously, diffusive velocities of the order of nanometres per second cannot describe the observed Al injection in nanowires. The complete map of  $v_D$  values fitting the measured  $k$  values is displayed in Fig. 4b. Note that the expression for  $\tau_{BL}$  can be



**Figure 4 | Calculations of partition coefficient and diffusive velocity.** **a**, The calculated partition coefficient,  $k$  (on colour scale and labelled on contours), for nanowires grown at 470 °C using a diffusive velocity ( $v_D$ ) of  $6.7 \text{ nm s}^{-1}$  obtained from an early solute trapping model<sup>25</sup>. Axes show  $f$  (given by  $\tau_{BL}/\tau$ ) and  $n$  (given by  $\tau_{BL}/\tau_e$ ); see text for details. Inset, the calculated  $k$  near  $(f;n) = (1;1)$ . **b**, The map in  $(f;n)$  space of diffusive velocities,  $v_D$  (on colour scale and labelled on contours), fitting the measured partition coefficient at 470 °C ( $k = 5.6 \times 10^{-2}$ ). It is important to indicate that only low values of  $f$  have a physical meaning in nanowire growth as the incubation time is much larger than the growth time of a single bilayer (that is,  $\tau_{BL} \ll \tau_1$ , corresponding to  $f \ll 1$ ). Note the logarithmic scales. In both **a** and **b**,  $v_{min}$  is taken as  $0.4 \text{ nm s}^{-1}$ , which is the average interface velocity corresponding to nanowires grown at 470 °C.

deduced from the model above:  $\tau_{BL} \approx (2R_{NW}/v_D)[(1-k)/(k-k_e)]$ . The extrapolation of the correlation between the equilibrium partition coefficient,  $k_e$ , and  $v_D$  obtained in laser resolidification<sup>25</sup> to our experimental conditions yields a diffusive velocity of  $\sim 2 \times 10^{11} \text{ nm s}^{-1}$  at 470 °C ( $k_e = 2.6 \times 10^{-6}$ ), corresponding to  $\tau_{BL} \approx 7 \text{ ns}$ . Interestingly, the time between two layers is  $\sim 0.75 \text{ s}$ , thus suggesting that the incubation time is very much larger than the time needed for the step to grow across the whole nanowire diameter, in qualitative agreement with *in situ* TEM observations<sup>8,26,29,30</sup>. Similarly, at the lowest growth temperature ( $k_e = 6.4 \times 10^{-7}$ ), the estimated  $\tau_{BL}$  is  $\sim 18 \text{ ns}$ . A reduction in growth temperature is anticipated to affect not only the rate of the thermally activated atomic jumps, but also other growth parameters that are critical for the incorporation of the catalyst atoms, as discussed below.

Using a step-flow growth model that takes catalyst dissolution into consideration, we derived an expression for the Al content in the nanowire as a function of growth parameters (Supplementary Information section G). Note, however, that the elucidation of the kinetically controlled atomic processes involved in the incorporation of Al would require a deep understanding of Al-induced Si surface reconstructions (Supplementary Information section H). From our analysis, it emerges that the second derivative of the Gibbs free energy of the liquid evaluated at the equilibrium composition,  $g_l''$ , and the characteristic supersaturation at which the nucleation occurs deterministically,  $\mu_c$ , are the two important parameters that influence catalyst injection. The quantity  $g_l''$  can be modified by adjusting Si solubility, whereas high  $\mu_c$  implies a high nucleation barrier. In principle, both parameters can be engineered through control of the catalyst chemistry (by adding impurities, for instance), thus providing the possibility of tailoring the properties of nanowires.

This flexibility in fabrication and functionalization of nanowires could be greatly enhanced owing to the availability of other catalysts that can act as dopants—for example, Bi (which yields n-type doping<sup>31,32</sup>) or Ga (which gives p-type doping<sup>33</sup>)—if their injection during the growth of Si or other group IV nanowires can be controlled. Our observations and predictions provide motivation to pursue the synthesis and characterization of atomically controlled nanowires with potentially adjustable morphology and physical characteristics that might offer new routes for catalytic assembly of nanowire-based devices.

## METHODS SUMMARY

The growth of Al-catalysed Si nanowires was accomplished by using monosilane,  $\text{SiH}_4$ , (diluted to 5% in argon) as a precursor in the temperature range  $\sim 400$ –470 °C (refs 10, 15). The partial pressure of the monosilane was held below 0.15 mbar during growth. Although the temperature used is below the Al-Si macroscopic eutectic temperature, the growth here is believed to be via the vapour–liquid–solid mechanism involving undercooled Al-Si nanodroplets. The morphology of Al-Si nanowires was characterized using an FEI dual-beam Nanolab 600 scanning electron microscope, a Philips CM 20T TEM operating at 200 kV, a JEOL JEM-4010 TEM operating at 400 kV, and an aberration-corrected FEI TITAN 80-300 analytical scanning TEM operating at 300 kV, which yields a spatial resolution of about 100 pm in both modes (TEM and scanning TEM). Additionally, this TEM is also equipped with an EDX detector having a detection limit of about 0.5 at.%. We use an ultraviolet laser-assisted LEAP (local electrode atom probe) tomograph (LEAP 4000XS, Cameca). The three-dimensional reconstructions and statistical analysis of concentration frequency were performed using Cameca's IVAS program. For the preparation of nanowire LEAP specimens, we have developed a focused ion beam (FIB)-based damage-free methodology to attach individual nanowires to commercially available Si microtips (Supplementary Information section B). The Si microtips were subsequently inserted into the LEAP tomograph's ultrahigh-vacuum chamber and cooled to 60 K before pulsed ultraviolet-laser-assisted evaporation analyses. To confirm that the preparation and laser-assisted evaporation do not affect the accuracy of the analysis, a reference sample consisting of an Al-capped Si needle was also analysed under exactly the same conditions, and the results demonstrate that the needle is Al-free. The binomial analysis was performed by sampling blocks of 200 atoms. The entire volume was first divided into columns along the  $z$  axis, with a cross-section in  $x$ - and  $y$ -directions, so that the volume  $dx dy dz$  contains 200 atoms on average. To obtain blocks with exactly 200 atoms, each column with cross-section  $dx dy$  was then cut into sections with  $dz$  as varied.

Received 9 August 2012; accepted 7 February 2013.

1. Yan, H. *et al.* Programmable nanowire circuits for nanoprocessors. *Nature* **470**, 240–244 (2011).
2. Kelzenberg, M. D. *et al.* Enhanced absorption and carrier collection in Si wire arrays for photovoltaic applications. *Nature Mater.* **9**, 239–244 (2010).
3. Hochbaum, A. & Yang, P. Semiconductor nanowires for energy conversion. *Chem. Rev.* **110**, 527–546 (2010).
4. Xie, P., Xiong, Q., Fang, Y., Qing, Q. & Lieber, C. M. Local electrical potential detection of DNA by nanowire-nanopore sensors. *Nature Nanotechnol.* **7**, 119–125 (2011).
5. Perea, D. E. *et al.* Direct measurement of dopant distribution in an individual vapor-liquid-solid nanowire. *Nature Nanotechnol.* **4**, 315–319 (2009).
6. Hoffmann, S. *et al.* Axial p-n junctions realized in silicon nanowires by ion implantation. *Nano Lett.* **9**, 1341–1344 (2009).
7. Ho, J. C. *et al.* Controlled nanoscale doping of semiconductors via molecular monolayers. *Nature Mater.* **7**, 62–67 (2008).
8. Wacaser, B. A. *et al.* Growth system, structure, and doping of aluminum-seeded epitaxial silicon nanowires. *Nano Lett.* **9**, 3296–3301 (2009).
9. Ke, Y. *et al.* Fabrication and electrical properties of Si nanowires synthesized by Al catalyzed vapor-liquid-solid growth. *Nano Lett.* **9**, 4494–4499 (2009).
10. Moutanabbir, O. *et al.* Atomically smooth p-doped silicon nanowires catalyzed by aluminum at low temperature. *ACS Nano* **5**, 1313–1320 (2011).
11. Choi, S.-Y., Fung, W. Y. & Lu, W. Growth and electrical properties of Al-catalyzed Si nanowires. *Appl. Phys. Lett.* **98**, 033108 (2011).
12. Trumbore, F. A. Solid solubilities of impurity elements in germanium and silicon. *Bell Syst. Tech. J.* **30**, 205–233 (1960).
13. Nast, O. & Wenham, S. R. Elucidation of the layer exchange mechanism in the formation of polycrystalline silicon by aluminum-induced crystallization. *J. Appl. Phys.* **88**, 124–132 (2000).
14. Civale, Y., Nanver, L. K., Alberici, S. G., Gammon, A. & Kelly, I. Accurate SIMS doping profiling of aluminum-doped solid-phase epitaxy silicon islands. *Electrochem. Solid-State Lett.* **11**, H74–H76 (2008).
15. Wang, Y., Schmidt, V., Senz, S. & Gösele, U. Epitaxial growth of silicon nanowires using aluminium catalyst. *Nature Nanotechnol.* **1**, 186–189 (2006).
16. Kim, B. J. *et al.* Kinetics of individual nucleation events observed in nanoscale vapor-liquid-solid growth. *Science* **322**, 1070–1073 (2008).
17. Seidman, D. N. & Stiller, K. An atom-probe tomography primer. *Mater. Res. Soc. Bull.* **34**, 717–724 (2009).
18. Moutanabbir, O., Isheim, D., Seidman, D. N., Kawamura, Y. & Itoh, K. M. Ultraviolet-laser atom-probe tomographic three-dimensional atom-by-atom mapping of isotopically modulated Si nanoscopic layers. *Appl. Phys. Lett.* **98**, 013111 (2011).
19. Murray, J. L. & McAlister, A. J. The Al-Si (aluminum-silicon) system. *Bull. Alloy Phase Diagrams* **5**, 74–84 (1984).
20. Hannon, J. B., Kodambaka, S., Ross, F. M. & Tromp, R. M. The influence of the surface migration of gold on the growth of silicon nanowires. *Nature* **440**, 69–71 (2006).
21. Trumbore, F. A. Solid solubilities of impurity elements in germanium and silicon. *Bell Syst. Tech. J.* **30**, 205–233 (1960).
22. Schwalbach, E. J. & Voorhees, P. W. Phase equilibrium in VLS-grown nanowires. *Nano Lett.* **8**, 3739–3745 (2008).
23. Baker, J. C. & Cahn, J. W. Solute trapping by rapid solidification. *Acta Metall.* **17**, 575–578 (1969).
24. White, C. W., Wilson, S. R., Appleton, B. R. & Young, F. W. Supersaturated substitutional alloys formed by ion implantation and pulsed laser annealing of group-III and group-V dopants in silicon. *J. Appl. Phys.* **51**, 738–749 (1980).
25. Reitano, R., Smith, P. M. & Aziz, M. J. Solute trapping of group III, IV, and V elements in silicon by an aperiodic stepwise growth mechanism. *J. Appl. Phys.* **76**, 1518–1529 (1994).
26. Wen, C.-Y., Tersoff, J., Reuter, M. C., Stach, E. A. & Ross, F. M. Step-flow kinetics in nanowire growth. *Phys. Rev. Lett.* **105**, 195502 (2010).
27. Aziz, M. J. M. J. Model for solute redistribution during rapid solidification. *J. Appl. Phys.* **53**, 1158–1168 (1982).
28. Goldman, L. M. & Aziz, M. J. Aperiodic stepwise growth model for the velocity and orientation dependence of solute trapping. *J. Mater. Res.* **2**, 524–527 (1987).
29. Wen, C.-Y. *et al.* Formation of compositionally abrupt axial heterojunctions in silicon-germanium nanowires. *Science* **326**, 1247–1250 (2009).
30. Kodambaka, S., Tersoff, J., Reuter, M. C. & Ross, F. M. Germanium nanowire growth below the eutectic temperature. *Science* **316**, 729–732 (2007).
31. Kumar, R., Narasimha Rao, K. & Phanib, A. R. Bismuth catalyzed growth of silicon nanowires by electron beam evaporation. *Mater. Lett.* **82**, 163–166 (2012).
32. Heitsch, A. T., Fanfair, D. D., Tuan, H. Y. & Korgel, B. A. solution–liquid–solid (SLS) growth of silicon nanowires. *J. Am. Chem. Soc.* **130**, 5436–5437 (2008).
33. Givargizov, E. I. & Sheftal, N. N. Morphology of silicon whiskers grown by the VLS-technique. *J. Cryst. Growth* **9**, 326–329 (1971).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** O.M. acknowledges support from NSERC, the Canada Research Chair, and École Polytechnique de Montréal (PIED). S.S. is grateful to Y. Wang for help in the growth of nanowires. The atom-probe tomography research was performed in the Northwestern University Center for Atom-Probe Tomography (NUCAPT), with partial support from the US-Israel Binational Science Foundation and the Max Planck Institute of Microstructure Physics.

**Author Contributions** O.M. led this research, analysed the data, developed the theoretical models, and wrote the manuscript. All authors commented on the manuscript. O.M., D.I. and D.N.S. designed the atom probe work. O.M. and S.S. discussed the initial experimental work. Nanowire growth was done using equipment maintained by S.S.. O.M. and H.B. established the FIB processing for LEAP specimens and H.B. conducted the FIB work. D.I. performed the atom probe analysis. E.P. did the TEM and EDX analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.M. (oussama.moutanabbir@polymtl.ca).

# July 2012 Greenland melt extent enhanced by low-level liquid clouds

R. Bennartz<sup>1</sup>, M. D. Shupe<sup>2</sup>, D. D. Turner<sup>3</sup>, V. P. Walden<sup>4</sup>, K. Steffen<sup>2,5</sup>, C. J. Cox<sup>4</sup>, M. S. Kulie<sup>6</sup>, N. B. Miller<sup>6</sup> & C. Pettersen<sup>6</sup>

Melting of the world's major ice sheets can affect human and environmental conditions by contributing to sea-level rise. In July 2012, an historically rare period of extended surface melting was observed across almost the entire Greenland ice sheet<sup>1,2</sup>, raising questions about the frequency and spatial extent of such events. Here we show that low-level clouds consisting of liquid water droplets ('liquid clouds'), via their radiative effects, played a key part in this melt event by increasing near-surface temperatures. We used a suite of surface-based observations<sup>3</sup>, remote sensing data, and a surface energy-balance model. At the critical surface melt time, the clouds were optically thick enough and low enough to enhance the downwelling infrared flux at the surface. At the same time they were optically thin enough to allow sufficient solar radiation to penetrate through them and raise surface temperatures above the melting point. Outside this narrow range in cloud optical thickness, the radiative contribution to the surface energy budget would have been diminished, and the spatial extent of this melting event would have been smaller. We further show that these thin, low-level liquid clouds occur frequently, both over Greenland and across the Arctic, being present around 30–50 per cent of the time<sup>3–6</sup>. Our results may help to explain the difficulties that global climate models have in simulating the Arctic surface energy budget<sup>7–9</sup>, particularly as models tend to under-predict the formation of optically thin liquid clouds at super-cooled temperatures<sup>6</sup>—a process potentially necessary to account fully for temperature feedbacks in a warming Arctic climate.

Over the past few decades, the Arctic has experienced warming amplified by a set of positive feedbacks; these feedbacks include increased sea ice melt, increased atmospheric water vapour and cloudiness, and changes in atmospheric circulation patterns<sup>10</sup>. This warming has resulted in an increased extent of surface melt of the Greenland ice sheet (GIS) observed by satellite since 1979<sup>11</sup>. July 2012 set a new record in melt extent, with melting observed over nearly the entire GIS<sup>1,2</sup>. At Summit Station, a brief melting period was observed on 11 July 2012 (Fig. 1). Ice-core records from the same location indicate such events occur only about once every 150 yr (ref. 1) on average, with the last occurring in 1889<sup>12</sup>. The July 2012 melt event was triggered by advection of unusually warm air, with temperatures at 500 m above ground only slightly below freezing (Fig. 1). However, the observed surface melt at Summit cannot be explained by warm air advection alone, because the surface temperature over the GIS is controlled by a balance of radiative and turbulent heat fluxes<sup>13–16</sup> that are sensitive to cloud and atmospheric properties.

The local energy balance at the surface can be understood by accounting for three forcings on surface temperature, namely, the net radiative flux divergence at the surface, heat exchange with the atmosphere, and heat exchange with the underlying ice. A simple parametric model of these processes can be written as:

$$\frac{\partial T_S}{\partial t} = \underbrace{\frac{\alpha+1}{c_p \rho H} \Delta F_{NET}}_{\text{Radiative forcing}} + \underbrace{\frac{T_a - T_S}{\tau_a}}_{\text{Atmospheric heat exchange}} + \underbrace{\frac{T_i - T_S}{\tau_i}}_{\text{Ice heat exchange}} \quad (1)$$

This model is used here to study the effect of clouds on the temporal development of surface temperature. In equation (1),  $T_S$  is the surface temperature,  $t$  is time,  $H$  is the height of the inversion layer,  $\alpha$  is a shape parameter characterizing the form of the temperature profile in the boundary layer,  $\rho$  is the average air density of the boundary layer,  $c_p$  is the specific heat of air,  $T_a$  is the temperature at the top of the inversion,  $T_i$  is the ice temperature at a depth low enough to be considered constant over the model integration time,  $\Delta F_{NET}$  is the net radiative flux divergence at the surface, and  $\tau_a$  and  $\tau_i$  are relaxation timescales for heat exchange processes in the atmosphere and ice, respectively. Whereas the two heat exchange terms are non-negligible, the radiative forcing plays a dominant role in the development of surface temperature<sup>14</sup>, which justifies the basic parameterizations of heat fluxes using relaxation times. In reality, atmospheric heat fluxes are complex functions of wind speed, turbulence state, temperature, and moisture profiles. The heat transfer into the surface depends on the snow's thermal conductivity but also on radiative heat exchange between different snow layers and solar radiation penetrating deeper into the snow layer<sup>17</sup>. The dependency of  $\tau_a$  on atmospheric stability is simply modelled by making  $\tau_a$  increase with increasing atmospheric stability, that is,  $\tau_a = f(T_a - T_S)$ . The model's response is not very sensitive to the choices of some of the model parameters; these parameters were fixed at  $H = 210$  m,  $\alpha = 2$ ,  $\tau_i = 15$  h, and  $T_i = -40$  °C. The particular choices of these parameters, as well as details of the model, are justified and outlined in the Supplementary Information.

The model assumes the temperature of the free atmosphere at the top of the inversion layer ( $T_a$ ) to be a lateral boundary condition determined by advection. This temperature was determined from 12-hourly radiosonde observations at Summit. The individual radiative fluxes constituting  $\Delta F_{NET}$ , the net radiative flux at the surface, are:

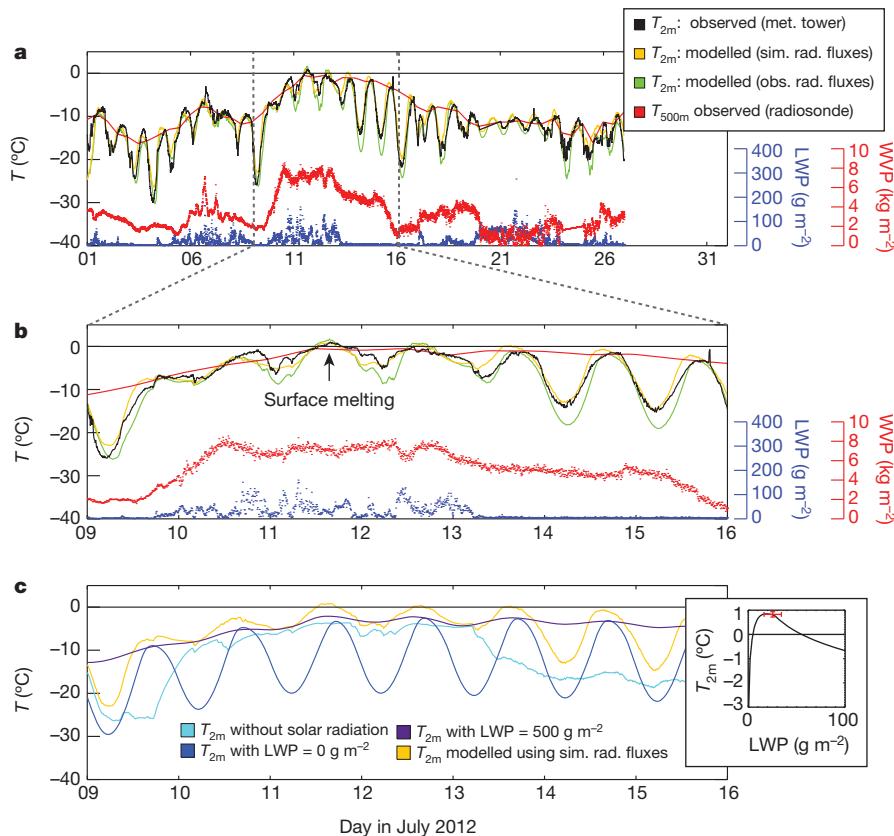
$$\Delta F_{NET} = F_{SW}^{\downarrow} - F_{SW}^{\uparrow} + F_{LW}^{\downarrow} - F_{LW}^{\uparrow} \quad (2)$$

The up and down arrows indicate the upwelling and downwelling radiation, respectively. The surface temperature model can be forced with observed or simulated radiative fluxes. Both approaches were performed for this study. For simulating fluxes, a simple atmospheric radiative transfer model was devised, which is outlined in the Supplementary Information. The shortwave component of the model ( $F_{SW}^{\uparrow,\downarrow}$ ) includes the radiative effects of water vapour, ozone, carbon dioxide, cloud liquid water, and surface albedo. The downwelling longwave component ( $F_{LW}^{\downarrow}$ ) accounts for thermal emission by atmospheric gases and cloud liquid water. The upwelling longwave radiation was simulated simply using Stefan–Boltzmann's law and a surface emissivity of unity, that is,  $F_{LW}^{\uparrow} = \sigma_{SB} T_S^4$ , where  $\sigma_{SB}$  is the Stefan–Boltzmann constant. The radiative transfer model does not account for ice clouds.

Figure 1a provides an overview of the temporal development of  $T_{2m}$ , the temperature at a height of two metres, both observed and simulated

<sup>1</sup>Atmospheric and Oceanic Sciences, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>2</sup>CIRES, University of Colorado and NOAA-ESRL, Boulder, Colorado 80305, USA. <sup>3</sup>NOAA-National Severe Storms Laboratory, Norman, Oklahoma 73072, USA. <sup>4</sup>Department of Geography, University of Idaho, Moscow, Idaho 83844, USA. <sup>5</sup>Swiss Federal Research Institute WSL, Birmensdorf, CH-8903, Switzerland.

<sup>6</sup>Space Science and Engineering Center, University of Wisconsin, Madison, Wisconsin 53706, USA.



**Figure 1 | Observed and simulated temporal evolution of the July 2012 surface melting event at Summit.** **a**, Temporal evolution of the temperature 2 m above the surface ( $T_{2m}$ ) for July 2012; **b**, expanded view of data from **a** for the extended GIS melting period (9–15 July 2012). In the top parts of **a** and **b**, the black curve shows  $T_{2m}$  observed by the NOAA meteorological (met.) tower, the red curve shows the temperature observed by radiosonde at 500 m above ground ( $T_{500m}$ ), the green curve shows the development of  $T_{2m}$  simulated using the surface energy balance model driven by observed radiative (obs. rad.) fluxes<sup>26</sup>, and the yellow curve shows the development of  $T_{2m}$  simulated using the surface energy balance model driven by simulated radiative (sim. rad.) fluxes based on water vapour path (WVP) and liquid water path (LWP). Values

of these two last quantities are shown in the lower parts of **a** and **b** with separate axes: blue (red) dots show LWP (WVP) values observed by an upward looking microwave radiometer. **c**, Model sensitivity studies with respect to clouds and solar radiation. Yellow line, as in **a** and **b**; turquoise curve,  $T_{2m}$  without solar radiation; blue curve,  $T_{2m}$  assuming a cloud-free atmosphere ( $LWP = 0 \text{ g m}^{-2}$ ); purple curve,  $T_{2m}$  assuming an atmosphere with a very thick cloud with constant LWP of  $500 \text{ g m}^{-2}$ . Inset, simulated maximum  $T_{2m}$  (corresponding to the position of the arrow ‘surface melting’ in **b**) as a function of an assumed constant LWP; red data point shows the observed  $T_{2m}$  and LWP at a time of  $\pm 30$  min around the time of maximum surface temperature (error bars,  $\pm 1$  s.d.).

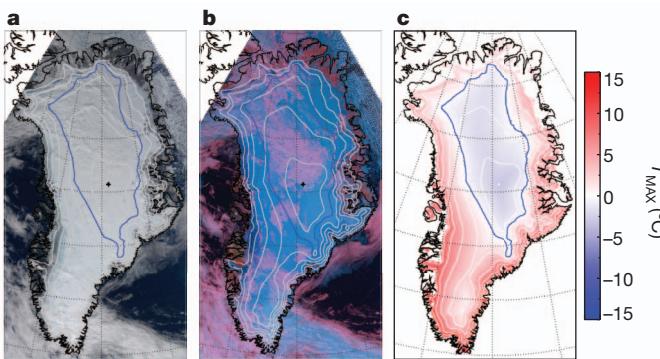
for July 2012 at Summit. For weaker surface inversions, this temperature will be very close to the surface temperature,  $T_{2m} \approx T_s$ . In the case of strong surface inversions,  $T_{2m}$  can be higher than  $T_s$ . Such cases might occur, for example, during the night under cloud-free conditions (for example, 14 July 2012). The cloud-free diurnal cycle of  $T_{2m}$  is of the order of  $10^\circ\text{C}$ , with low temperatures during night caused by less incoming solar radiation and strong longwave energy loss. (We note that the solar zenith angle at the beginning of July at Summit is about  $85^\circ$  at local midnight and around  $52^\circ$  at local noon.) The surface melting period on 11 July 2012 was marked by low cloud bases, which occur very frequently over Summit. Low, liquid-bearing clouds were also present on 12 July 2012, followed by three cloud-free days. During the cloud-free period, and in particular on 13 July, warm air was still present over Summit but surface temperatures did not rise above  $0^\circ\text{C}$ . A visual overview of cloudiness for July 2012 (as well as July 2010 and 2011 for comparison) is provided in Supplementary Fig. 5 and Supplementary Table 2, based on radar and lidar observations.

Despite its simplicity, the prognostic surface temperature model described in equation (1) reasonably captures the long-term variability, the phase and amplitude of the diurnal cycle, as well as its modulation by liquid clouds. Deviations occur in the presence of high ice clouds, which were not included in the simulations. Note that modelling results based on surface radiative flux observations agree well with the results where the surface radiative fluxes were computed using

cloud observations and the radiative transfer model. This agreement confirms that ice clouds play a minor role in this particular study because the simulated radiative fluxes do not include effects of ice clouds whereas the observed fluxes do. The model also predicts the timing and strength of the surface melt event in July 2012 with reasonable accuracy (see Fig. 1b).

The modulating role of low-level liquid-containing clouds on the surface energy balance can be understood by considering two competing effects. First, these clouds reflect solar radiation to space, reducing shortwave energy available for surface warming and dampening the diurnal cycle. Second, low-level clouds radiate energy downwards in the infrared. This process is efficient even for very thin clouds. At liquid water path (LWP) values greater than  $20 \text{ g m}^{-2}$ , clouds become nearly completely opaque in the infrared<sup>18,19</sup> in which case the downwelling longwave radiation is determined almost entirely by cloud temperature. For the dry atmosphere observed over the GIS, cloudy-sky downwelling longwave fluxes can easily be  $100 \text{ W m}^{-2}$  higher than those measured for cloud-free conditions.

The relative balance of decreased solar radiation and increased downward longwave radiation depends on cloud optical properties, which are most strongly modulated by variations in LWP. Figure 1c shows model sensitivity studies for two extreme cases: no cloud (that is,  $LWP = 0 \text{ g m}^{-2}$ ) and an extremely thick cloud with LWP constant at  $500 \text{ g m}^{-2}$ . In the cloud-free case, a strong diurnal cycle exists but



**Figure 2 | Observed spatial distribution of clouds over Greenland on 11 July 2012 and their effect on surface temperature.** **a**, Near-true-colour satellite image of Greenland observed by MODIS (on the Aqua satellite) on 11 July 2012, 14:55–15:05 UTC (MODIS channels 1, 4, 3). **b**, False-colour image for the same time period highlighting liquid-bearing clouds in purple (MODIS channels 7, 2, 1). The star shows the position of Summit. White contour lines give terrain height in 500-m intervals starting at 1,000 m. **c**, The maximum simulated temperature under cloud-free conditions (see text for details). The blue line corresponds to the 0 °C isothermal—that is, the height below which melting would occur under cloud-free conditions. This line is also shown in **a** and **b** as a reference.

near-surface temperatures do not rise above  $-4^{\circ}\text{C}$ . For the thick cloud, near-surface temperature follows the cloud temperature but does not exceed a value of  $-3^{\circ}\text{C}$ .

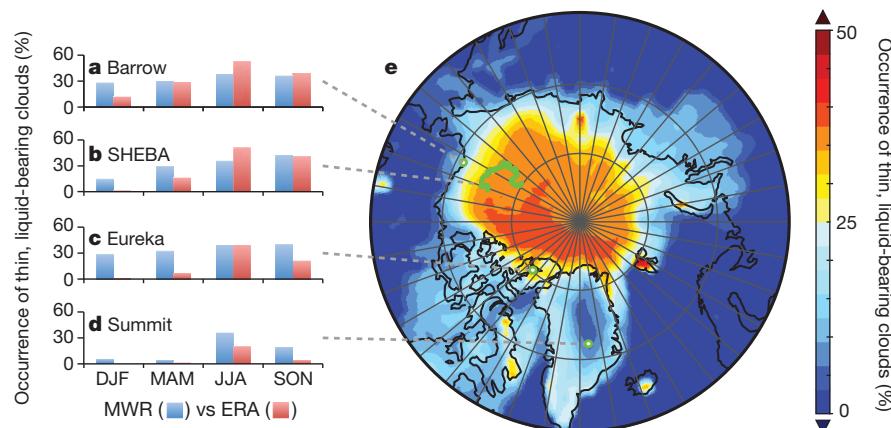
In neither of these two extreme cases does the near-surface temperature rise to values above the melting point. Only for a limited range of intermediate LWP values do the combined longwave and shortwave radiative effects conflate to push temperatures above  $0^{\circ}\text{C}$ . Indeed, the energy balance model indicates that surface melting only occurs in this case for LWP values between  $10$  and  $40 \text{ g m}^{-2}$  (see Fig. 1c inset). Within this range, the cloud is opaque enough to ‘trap’ longwave radiation but still sufficiently thin to allow enough solar radiation to penetrate through to the surface. Clouds of this nature probably affected broad spatial areas of the GIS during this event. Figure 2a and b provide an overview of the cloud situation over the GIS on 11 July 2012, observed by the Moderate Resolution Imaging Spectrometer

(MODIS) on NASA’s Aqua satellite, and show extended liquid clouds over large parts of the central GIS.

Surface melting would not have occurred over Summit in July 2012 without the optically thin, liquid-water-containing clouds. However, temperatures were warm enough during this time to trigger melting without clouds at lower-elevation regions of the GIS. The diurnal cycle and the total amount of incoming solar radiation also vary with latitude. A second sensitivity study was carried out to estimate the relative impact of these two factors—that is, terrain height above sea level and latitude—on surface melting. Figure 2c shows the result of this sensitivity study, assuming cloud-free conditions and adjusting the model (equation (1)) for height effects on atmospheric temperature, down-welling longwave radiation, and incoming solar radiation as function of latitude. For all temperature dependencies on surface height, a climatological mean temperature lapse rate of  $7.1 \text{ K km}^{-1}$  was used<sup>20</sup>. Terrain slope and orientation were not considered, as terrain slopes on top of the GIS are generally small. This sensitivity study is not expected to reproduce the exact conditions observed over Greenland, but it does provide insight into the interplay of terrain height, available solar radiation, and clouds. The maximum simulated temperature (Fig. 2c) did not exceed  $0^{\circ}\text{C}$  at terrain heights above roughly 2,700 m for regions south of Summit. At more northerly locations, where less solar radiation is available, these terrain heights fell below 2,000 m (see the blue  $0^{\circ}\text{C}$  isothermal in all three panels of Fig. 2). In comparison, for a homogeneous cloud of  $30 \text{ g m}^{-2}$ , the model predicted melting everywhere (not shown). For this example model exercise, the spatial extent of GIS melt conditions increased from 61% to 100% as a result of the optically thin clouds.

This study highlights the intricate role of clouds in modulating the energy balance at the surface above the GIS in boreal summer. Thin, boundary-layer, liquid-containing clouds can change the surface energy balance towards warmer or colder surface temperatures depending on cloud optical thickness. In the case examined here, these clouds played a critical role in enhanced surface melt at Summit, Greenland, and probably contributed strongly to the 97% spatial melt extent observed over the GIS in early July 2012<sup>1</sup>.

Thin, liquid-bearing clouds are ubiquitous across the Arctic<sup>6</sup>. Ground-based observations and model reanalysis products from the European Center for Medium Range Weather Forecasting (ERA-Interim) show occurrence frequencies ranging from 20% to more than



**Figure 3 | Frequency of occurrence of thin, liquid-bearing clouds.** For the purpose of this plot, ‘thin, liquid-bearing’ clouds are defined as clouds in the range of  $10 \text{ g m}^{-2} < \text{LWP} < 60 \text{ g m}^{-2}$ , corresponding to the range of maximum enhanced cloud radiative forcing at the surface. **a–d**, Comparisons of ground-based observed (blue, microwave radiometer (MWR)) and ERA-Interim simulated (red, ERA) frequencies of occurrence of these clouds for four Arctic observation sites for all seasons; **a**, Barrow, Alaska; **b**, Surface Heat Budget of the Arctic Ocean (SHEBA) experiment (a research vessel frozen in

the pack ice to study the ocean and atmosphere); **c**, Eureka, Nunavut (research base on Ellesmere Island, Canada); and **d**, Summit, Greenland. DJF, December–February; MAM, March–May; JJA, June–August; SON, September–November. **e**, Circumpolar map of the frequency of occurrence of these clouds from 32 yr of ERA reanalysis (1979–2011). The plot in **e** is conditionally sampled to only include cases with solar zenith angle lower than  $80^{\circ}$  and a surface albedo higher than 0.5; locations of the observation sites (**a–d**) are indicated.

50% of the time for four observational sites in summer and for most observational sites also in spring and autumn (Fig. 3a–d). (See Supplementary Information section 6 for details on the temporal coverage of the different observation sites; please also note that ERA-Interim histograms shown in Fig. 3a–d are restricted to the time period when observations are available.) Based on ERA-Interim reanalysis, Fig. 3e shows the occurrence of thin, liquid-bearing clouds in situations where the surface albedo is larger than 0.5 and at least some solar radiation is available (solar zenith angle less than 80°). These situations are most conducive to enhanced surface warming by the combined effects of solar and infrared warming discussed above<sup>21–24</sup>. Thin liquid clouds occur most frequently over sea-ice-covered areas north of 80°N, as well as over the Beaufort Sea and the east Siberian Sea, highlighting the importance of such clouds not only for the GIS but also for the surface energy balance over sea ice.

To simulate the surface energy balance correctly, it is imperative for climate models to accurately represent these liquid-containing clouds and their occurrence fraction, radiative properties and responses to a changing Arctic climate system. The comparisons provided in Fig. 3a–d highlight a significant shortcoming in our current modelling capabilities of Arctic clouds. ERA reanalysis provides mostly reasonable results for thin, liquid-cloud occurrence in summer months. However, the model significantly underestimates the occurrence of such clouds in winter and spring for most sites and for Summit in all seasons. A similar underestimation of the occurrence of thin, liquid clouds was recently reported<sup>6</sup> for global climate models based on comparisons with satellite-derived estimates of cloudiness. The discrepancies were attributed largely to the simplified treatment of cloud phase in models, which typically does not allow for liquid clouds to form at temperatures lower than about –20 °C. A similar phase determination scheme is employed in the ERA-Interim data, potentially explaining the significant low bias in modelled liquid-cloud occurrence over Summit (Fig. 3d). These shortcomings bias the Arctic surface energy balance predicted by present-day weather and climate models<sup>6</sup>, and limit the ability of these models to predict the cloud response to Arctic climate change and possible feedbacks.

## METHODS SUMMARY

Ground-based infrared, microwave, radar and lidar remote sensing observations, as well as radiosonde data, were collected as part of the project 'Integrated characterization of energy, clouds, atmospheric state, and precipitation at Summit' (ICECAPS)<sup>3</sup>. Cloud liquid-water path estimates were derived from multi-channel microwave measurements<sup>4</sup>. Downwelling longwave fluxes were derived from high-resolution infrared observations and cloud-free fluxes were calculated on the basis of radiosonde observations<sup>25</sup>. In addition, shortwave and longwave upwelling and downwelling radiative fluxes were observed from a suite of broadband radiometers<sup>26,27</sup>, while temperatures at a height of two metres were observed from a meteorological tower operated by NOAA. All Summit, SHEBA, and Eureka ground-based data are available from the NOAA ESRL site (<ftp://ftp1.esrl.noaa.gov/psd3/arctic/>). Barrow data are available from the DOE ARM archive (<http://www.archive.arm.gov/>). MODIS satellite observations were obtained from NASA's Level 1 and Atmosphere Archive and Distribution System. ECMWF ERA Reanalysis data are available from the ECMWF data server (<http://www.ecmwf.int>).

Received 16 October 2012; accepted 8 February 2013.

- Buis, A. & Cole, S. Satellites see unprecedented Greenland Ice Sheet melt. (NASA/JPL press release, 24 July 2012); available at <http://www.jpl.nasa.gov/news/news.cfm?release=2012-217> (2012).
- Nghiem, S. V. et al. The extreme melt across the Greenland Ice Sheet in 2012. *Geophys. Res. Lett.* **39**, L20502, <http://dx.doi.org/10.1029/2012GL053611> (2012).
- Shupe, M. D. et al. High and dry: new observations of tropospheric and cloud properties above the Greenland Ice Sheet. *Bull. Am. Meteorol. Soc.* **94**, 169–186 (2013).
- Turner, D. D. et al. Retrieving liquid water path and precipitable water vapor from the atmospheric radiation measurement (ARM) microwave radiometers. *IEEE Trans. Geosci. Remote Sensing* **45**, 3680–3690 (2007).
- Shupe, M. D. et al. Clouds at Arctic atmospheric observatories. Part I: occurrence and macrophysical properties. *J. Appl. Meteorol. Climatol.* **50**, 626–644 (2011).
- Cesana, G., Kay, J. E., Chepfer, H., English, J. M. & de Boer, G. Ubiquitous low-level liquid-containing Arctic clouds: new observations and climate model constraints

- from CALIPSO-GOCCP. *Geophys. Res. Lett.* **39**, L20804, <http://dx.doi.org/10.1029/2012GL053385> (2012).
- de Boer, G. et al. A characterization of the present-day Arctic atmosphere in CCSM4. *J. Clim.* **25**, 2676–2695 (2012).
  - Klein, S. A. et al. Intercomparison of model simulations of mixed-phase clouds observed during the ARM mixed-phase Arctic cloud experiment. I: Single-layer cloud. *Q. J. R. Meteorol. Soc.* **135**, 979–1002 (2009).
  - Sandvik, A., Biryulina, M., Kvamstø, N. G., Stammes, J. J. & Stammes, K. Observed and simulated microphysical composition of arctic clouds: data properties and model validation. *J. Geophys. Res.* **112**, D05205, <http://dx.doi.org/10.1029/2006JD007351> (2007).
  - Serreze, M. C. & Barry, R. G. Processes and impacts of Arctic amplification: A research synthesis. *Global Planet. Change* **77**, 85–96 (2011).
  - Mernild, S. H., Mote, T. & Liston, G. E. Greenland ice sheet surface melt extent and trends: 1960–2010. *J. Glaciol.* **57**, 621–628 (2011).
  - Clausen, H. B., Gundestrup, N. S., Johnson, S. J., Bindschadler, R. & Zwally, J. Glaciological investigations in the Crete area, central Greenland: a search for a new deep drilling site. *Ann. Glaciol.* **10**, 10–15 (1988).
  - Ohmura, A. Physical basis for the temperature-based melt-index method. *J. Appl. Meteorol.* **40**, 753–761 (2001).
  - Hudson, S. R. & Brandt, R. E. A look at the surface-based temperature inversion on the Antarctic plateau. *J. Clim.* **18**, 1673–1696 (2005).
  - Brandt, R. E. & Warren, S. G. Temperature measurements and heat transfer in near-surface snow at the South Pole. *J. Glaciol.* **43**, 339–351 (1997).
  - Bintanja, R. & van den Broeke, M. The influence of clouds on the radiation budget of ice and snow surfaces in Antarctica and Greenland in summer. *Int. J. Climatol.* **16**, 1281–1296 (1996).
  - Brandt, R. E. & Warren, S. G. Solar-heating rates and temperature profiles in Antarctic snow and ice. *J. Glaciol.* **39**, 99–110 (1993).
  - Turner, D. D. Improved ground-based liquid water path retrievals using a combined infrared and microwave approach. *J. Geophys. Res.* **112**, D15204, <http://dx.doi.org/10.1029/2007JD008530> (2007).
  - Turner, D. D. et al. Thin liquid water clouds — their importance and our challenge. *Bull. Am. Meteorol. Soc.* **88**, 177–190 (2007).
  - Steffen, K. & Box, J. Surface climatology of the Greenland ice sheet: Greenland climate network 1995–1999. *J. Geophys. Res. D* **106**, 33951–33964 (2001).
  - Intrieri, J. M. et al. An annual cycle of Arctic surface cloud forcing at SHEBA. *J. Geophys. Res.* **107**, 8039, <http://dx.doi.org/10.1029/2000JC000439> (2002).
  - Shupe, M. D. & Intrieri, J. M. Cloud radiative forcing of the Arctic surface: the influence of cloud properties, surface albedo, and solar zenith angle. *J. Clim.* **17**, 616–628 (2004).
  - Pavolonis, M. J. & Key, J. R. Antarctic cloud radiative forcing at the surface estimated from the AVHRR Polar Pathfinder and ISCCP D1 datasets, 1985–93. *J. Appl. Meteorol.* **42**, 827–840 (2003).
  - Persson, P. O. G. Onset and end of the summer melt season over sea ice: thermal structure and surface energy perspective from SHEBA. *Clim. Dyn.* **39**, 1349–1371 (2012).
  - Cox, C. J., Walden, V. P. & Rowe, P. M. A comparison of the atmospheric conditions at Eureka, Canada, and Barrow, Alaska (2006–2008). *J. Geophys. Res.* **117**, D12204, <http://dx.doi.org/10.1029/2011JD017164> (2012).
  - Ohmura, A. et al. Baseline Surface Radiation Network (BSRN/WCRP): new precision radiometry for climate research. *Bull. Am. Meteorol. Soc.* **79**, 2115–2136 (1998).
  - Munneke, P. K. et al. The role of radiation penetration in the energy budget of the snowpack at Summit, Greenland. *Cryosphere* **3**, 277–306 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** ICECAPS is supported by the US National Science Foundation (grants ARC-0904152, 0856773 and 0856559) as part of the Arctic Observing Network (AON) programme. Two-metre temperature and radar measurements were provided by the NOAA Earth System Research Laboratory. Lidar and ceilometer measurements were provided by the Department of Energy Atmospheric Radiation Measurement programme. We thank the extended ICECAPS team of scientists, engineers, students and field technicians for obtaining high-quality atmosphere and cloud observations at Summit.

**Author Contributions** R.B. conceived this study, developed the surface temperature model, and performed most of the data analysis; M.D.S. coordinated ICECAPS measurement streams and contributed to interpretation of cloud-surface interactions and Arctic cloud context; D.D.T. performed the physical retrievals to derive precipitable water vapour and liquid water path from the microwave radiometers, and analysed the cloud height distributions over Summit for the three July periods; V.P.W. served as Principal Investigator for the ICECAPS project, and oversaw the calculation of longwave downwelling clear-sky and all-sky fluxes; K.S. provided the broadband radiative flux observations; C.J.C. calculated longwave downwelling clear-sky fluxes from radiosonde observations and all-sky fluxes from spectrally highly resolving infrared observations; M.S.K. provided support developing multi-frequency microwave radiometer retrievals and retrieving/analysing operational datasets; N.B.M. provided ceilometer derived cloud fraction calculations and collected various operational datasets; and C.P. helped operate the radar, and helped identify the MODIS data and other observational datasets.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (bennartz@aos.wisc.edu).

# Draft genome of the wheat A-genome progenitor *Triticum urartu*

Hong-Qing Ling<sup>1\*</sup>, Shancen Zhao<sup>2,3\*</sup>, Dongcheng Liu<sup>1\*</sup>, Junyi Wang<sup>1,2\*</sup>, Hua Sun<sup>1\*</sup>, Chi Zhang<sup>2\*</sup>, Huajie Fan<sup>1</sup>, Dong Li<sup>2</sup>, Lingli Dong<sup>1</sup>, Yong Tao<sup>2</sup>, Chuan Gao<sup>2</sup>, Huilan Wu<sup>1</sup>, Yiwen Li<sup>1</sup>, Yan Cui<sup>1</sup>, Xiaosen Guo<sup>2</sup>, Shusong Zheng<sup>1</sup>, Biao Wang<sup>1</sup>, Kang Yu<sup>1</sup>, Qinsi Liang<sup>2</sup>, Wenlong Yang<sup>1</sup>, Xueyuan Lou<sup>1</sup>, Jie Chen<sup>2</sup>, Mingji Feng<sup>2</sup>, Jianbo Jian<sup>2</sup>, Xiaofei Zhang<sup>1</sup>, Guangbin Luo<sup>1</sup>, Ying Jiang<sup>2</sup>, Junjie Liu<sup>2</sup>, Zhaobao Wang<sup>2</sup>, Yuhui Sha<sup>2</sup>, Bairu Zhang<sup>1</sup>, Huajun Wu<sup>4</sup>, Dingzhong Tang<sup>1</sup>, Qianhua Shen<sup>1</sup>, Pengya Xue<sup>1</sup>, Shenhao Zou<sup>1</sup>, Xiujie Wang<sup>4</sup>, Xin Liu<sup>1</sup>, Famin Wang<sup>1</sup>, Yaping Yang<sup>1</sup>, Xueli An<sup>1</sup>, Zhenying Dong<sup>1</sup>, Kunpu Zhang<sup>1</sup>, Xiangqi Zhang<sup>1</sup>, Ming-Cheng Luo<sup>5</sup>, Jan Dvorak<sup>5</sup>, Yiping Tong<sup>4</sup>, Jian Wang<sup>2</sup>, Huanming Yang<sup>2</sup>, Zhensheng Li<sup>1</sup>, Daowen Wang<sup>1</sup>, Aimin Zhang<sup>1</sup> & Jun Wang<sup>2,6,7</sup>

Bread wheat (*Triticum aestivum*, AABBDD) is one of the most widely cultivated and consumed food crops in the world. However, the complex polyploid nature of its genome makes genetic and functional analyses extremely challenging. The A genome, as a basic genome of bread wheat and other polyploid wheats, for example, *T. turgidum* (AABB), *T. timopheevii* (AAGG) and *T. zhukovskiyi* (AAGGA<sup>m</sup>A<sup>m</sup>), is central to wheat evolution, domestication and genetic improvement<sup>1</sup>. The progenitor species of the A genome is the diploid wild einkorn wheat *T. urartu*<sup>2</sup>, which resembles cultivated wheat more extensively than do *Aegilops speltoides* (the ancestor of the B genome<sup>3</sup>) and *Ae. tauschii* (the donor of the D genome<sup>4</sup>), especially in the morphology and development of spike and seed. Here we present the generation, assembly and analysis of a whole-genome shotgun draft sequence of the *T. urartu* genome. We identified protein-coding gene models, performed genome structure analyses and assessed its utility for analysing agronomically important genes and for developing molecular markers. Our *T. urartu* genome assembly provides a diploid reference for analysis of polyploid wheat genomes and is a valuable resource for the genetic improvement of wheat.

Bread wheat is one of the most important food crops worldwide, and provides about 20% of the calories consumed by humans<sup>5</sup>. To accelerate wheat improvement, a substantial amount of research has been conducted on the genome. The International Wheat Genome Sequencing Consortium aims at flow-sorting and sequencing the individual chromosomes of bread wheat, and significant progress has been made with several chromosomes, for example 3B (ref. 6) and 4A (ref. 7). More recently, a whole-genome shotgun sequence analysis of bread wheat and its diploid relatives<sup>8</sup> has allocated more than 60% of the genes to the A, B and D genomes with more than 70% confidence. The sequence of diploid progenitor genomes will allow the complete and unambiguous assignment of their homeologous relationships.

We sequenced *T. urartu* accession G1812 (PI428198) using a whole-genome shotgun strategy on the Illumina HiSequation (2000) platform, and assembled the genome using SOAPdenovo (v. 1.05)<sup>9</sup> with 448.49 gigabases (Gb) of filtered high-quality sequence data (Supplementary Information). We estimated the genome size of *T. urartu* to be 4.94 Gb (Supplementary Information), which is consistent with previous reports of 4.8–5.7 Gb (refs 10, 11). The genome assembly reached 3.92 Gb with a contig N50 size (at which 50% of assembly was covered) of 3.42 kilobases (kb). After gap closure, the draft assembly was 4.66 Gb with a scaffold N50 length of 63.69 kb (Table 1 and Supplementary Information). The length of the contigs that contained intact or partial genes ranged from

200 base pairs (bp) to 65.8 kb, with an average length of 9.91 kb. The assembly was evaluated by comparisons with published bacterial artificial chromosome and expressed sequence tag (EST) sequences and by validation with PCR (Supplementary Information), and both indicated that the draft sequence had extensive genome coverage with high accuracy. The distribution of GC content in the *T. urartu* genome was comparable with those in the genomes of rice<sup>12</sup>, maize<sup>13</sup>, sorghum<sup>14</sup> and *Brachypodium distachyon*<sup>15</sup> (Supplementary Information).

Genome annotation of the assembly was performed as described in Supplementary Information. About 66.88% of the *T. urartu* assembly was identified as repetitive elements, including long terminal repeat retrotransposons (49.07%), DNA transposons (9.77%) and unclassified elements (8.04%) (Supplementary Information). The proportion of repetitive DNA was lower than the roughly 80% previously reported<sup>16</sup>, which is probably due to a decreased incorporation of repeat sequence reads into the assemblies.

To facilitate gene prediction, we generated a 116.65-megabase (Mb) transcriptome of *T. urartu* with 67.14 Gb of RNA-Seq data from eight different tissues and treatments using the HiSequation (2000) platform, and 49,935 assembled transcripts from six tissues using the Roche 454 sequencing platform (Supplementary Information). These data, together with publicly available ESTs from hexaploid wheat, and homologues from sequenced grass genomes<sup>12–15,17</sup>, were used as evidence in gene prediction (Supplementary Information). In total, we predicted 34,879 protein-coding gene models. The average gene size was 3,207 bp, with a mean of 4.7 exons per gene, which was similar to that found for *B. distachyon* (5.2)<sup>15</sup> but slightly higher than that of rice (3.8)<sup>12</sup>, maize (4.1)<sup>13</sup> and sorghum (4.3)<sup>14</sup>. In comparison with the 28,000 genes estimated for the A genome of hexaploid wheat<sup>7</sup>, our gene set for *T. urartu* contained 6,800 more members, indicating a more complete representation of genes in our analysis. However, the different approaches used in this work and in a previous study<sup>7</sup>, and the extensive loss of genes in the hexaploid A genome compared with its diploid progenitor<sup>8</sup>, may also have contributed to this difference.

We also obtained 14,222,170 small RNA (sRNA) reads (18–30 bp) representing 4,369,970 unique sRNA tags. In total, 412 conserved and 24 new microRNAs (miRNAs) distributed into 116 families were identified. Comparison with the miRNAs of five monocots and five dicots showed that 73 miRNA families were specific to monocots, of which 23 were uniquely present in *T. urartu*. We predicted 244 target genes for these miRNAs and found that the target gene (TRIUR3\_06170) of miRNA MIR5050 responded to cold treatment, which provides a

<sup>1</sup>State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. <sup>2</sup>BGI-Shenzhen, Shenzhen 518038, China. <sup>3</sup>State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Hong Kong. <sup>4</sup>State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. <sup>5</sup>Department of Plant Sciences, University of California, Davis, California 95616, USA. <sup>6</sup>Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. <sup>7</sup>King Abdulaziz University, Jeddah 21589, Saudi Arabia.

\*These authors contributed equally to this work.

**Table 1 | Sequencing and draft genome assembly statistics**

| Assembly step         | DNA library insert size (bp) | Effective data (Mb) | N50 (bp) | Total length (Mb) | Nst† (Mb) | Number >2 kb |
|-----------------------|------------------------------|---------------------|----------|-------------------|-----------|--------------|
| Contig construction   | 167–796                      | 357,810             | 3,422    | 3,920             | —         | 385,430      |
| Scaffold construction | 167–20,000                   | 90,682              | 64,532   | —                 | 1,258     | —            |
| Gap closure*          | 167–796                      | 357,810             | 63,687   | 4,660             | 738       | 81,689       |

\* Final statistics of scaffolds after gap closure.

† The added N number during scaffolding.

new resource for investigating the regulation of cold adaptation through miRNA (Supplementary Information).

The gene families of *T. urartu* were compared with those of rice<sup>12</sup>, maize<sup>13</sup>, sorghum<sup>14</sup> and *B. distachyon*<sup>15</sup> using OrthoMCL<sup>18</sup> (Supplementary Information). We identified 24,339 families in the five grasses. Of these, 9,836 families, which contained 68,464 genes, were

common to all five species. Another 1,103 families, containing 3,425 genes, were specific to *T. urartu* (Fig. 1a). GO analysis of the *T. urartu*-specific families revealed that 556, 230 and 841 genes were involved in biological processes, cellular compounds and molecular functions, respectively. In total, 2,067 Pfam domains were shared among the five species. Of these, 14 Pfam domains had differences in member numbers in *T. urartu* compared with the other four grasses (Fig. 1b). These included NB-ARC and serine-threonine/tyrosine-protein kinase domains that were markedly increased in *T. urartu*, and C3HC4 RING-type and pathogenesis-related transcriptional factor/ERF domains that were significantly decreased. However, determination of the significance and accuracy of these differences will require more detailed analysis.

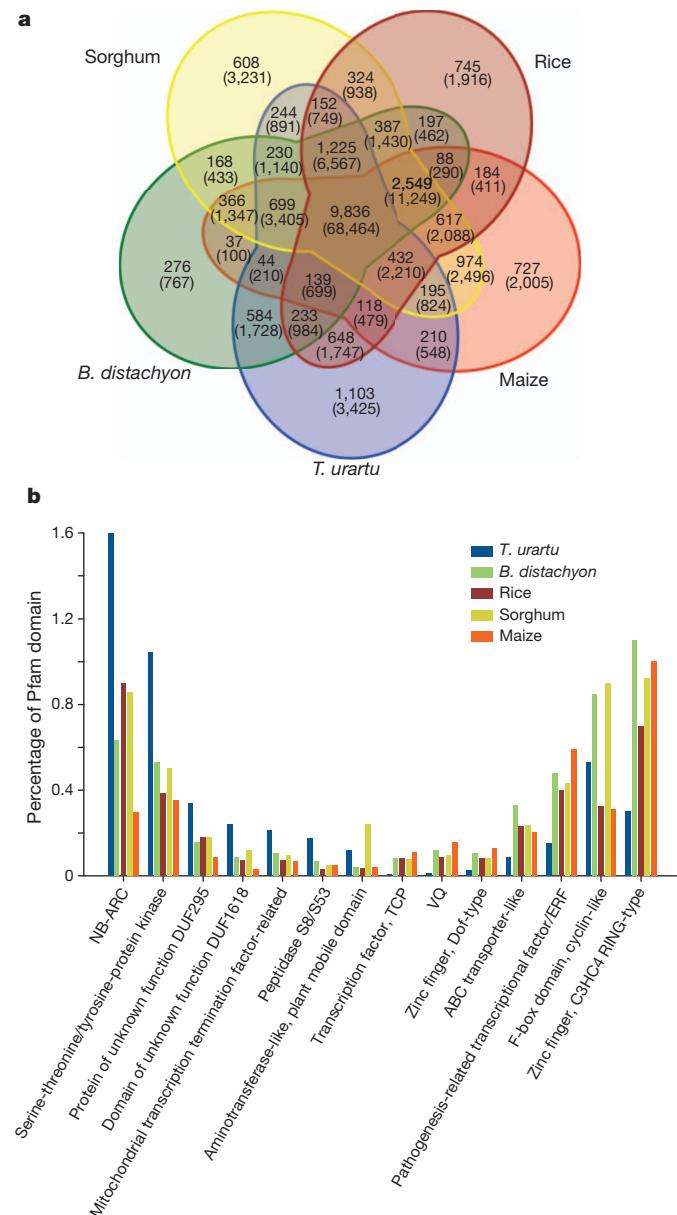
Given that NB-ARC domain proteins function mainly in disease resistance<sup>19</sup>, we analysed the genes encoding R proteins in the *T. urartu* genome and identified 593 such genes, which were more abundant than in *B. distachyon* (197), rice (460), maize (106) and sorghum (211) (Supplementary Information). In contrast with barley genome data<sup>17</sup>, the ratio of NBS-LRR type genes in *T. urartu* (1.21%) was also substantially higher than that in barley (0.73%). These analyses indicate that there was a specific expansion of R genes in the *T. urartu* genome.

The scaffolds and gene models of *T. urartu* were assigned to chromosomes by using genetically mapped bread wheat ESTs<sup>20</sup> as queries to search for homologous sequences in the *T. urartu* assembly (Supplementary Information). A total of 8,715 scaffolds, harbouring 14,578 genes (41.8% of the total predicted genes) were mapped to 45 chromosomal regions of the wheat A genome. Syntenic alignments between the *T. urartu* and *B. distachyon* genomes were constructed by using a set of 14,578 orthologous genes (Fig. 2a). These gene-based alignments conform, and supply more details, to the broad framework of genome synteny between wheat and *B. distachyon* proposed previously<sup>15</sup>.

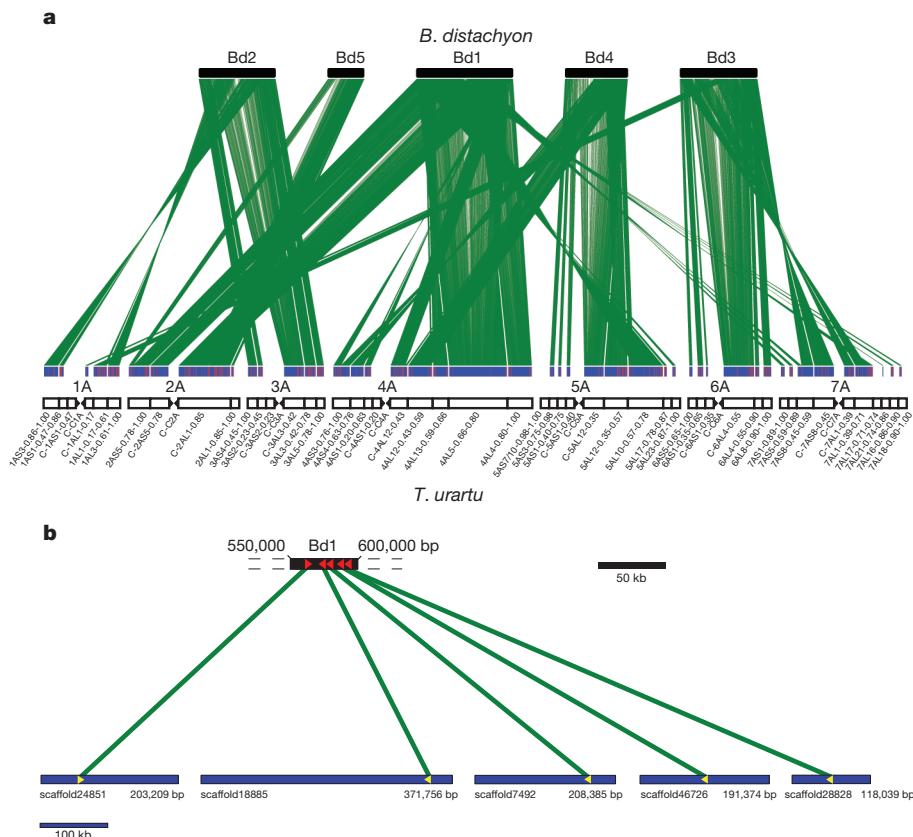
The 4.94-Gb *T. urartu* genome is more than 18 times larger than the 272-Mb genome of *B. distachyon*. Given that the average gene size of *T. urartu* is similar to that of *B. distachyon*, and the predicted gene number (34,879) is only about 1.37-fold that of *B. distachyon* (25,532), the larger genome size of *T. urartu* might be due to increased intergenic spaces. We therefore compared the intergenic space of the syntenic blocks between *T. urartu* and *B. distachyon* (Supplementary Information). About 21% of *T. urartu* genes had similarly sized intergenic spaces to those in *B. distachyon*, but most *T. urartu* genes were separated by a greatly increased intergenic space enriched in *Gypsy* and *Copia* retrotransposons, and were present in separate scaffolds (Fig. 2b). This provides the genome sequence-scale evidence for the role of repeat expansion in genome size enlargement during the evolution of the tribe Triticeae.

We next demonstrated the utility of the *T. urartu* draft genome sequence for finding agronomically important genes through identifying the *T. urartu* homologue of *OsGASR7*, a gibberellin-regulated gene that controls grain length in rice<sup>21</sup>. We found two haplotypes (H1 and H2) for *TuGASR7* in 92 diverse *T. urartu* accessions collected from different regions. H1 was significantly associated with greater values of grain length and grain weight (Fig. 3 and Supplementary Information). We also found natural variation of *TuGASR7* in bread wheat, with the elite variant associated with improved yield-related traits, suggesting that *TuGASR7* is of use for the improvement of wheat yield (Supplementary Information).

The *T. urartu* assembly also served as a rich resource for the development of genetic markers for molecular breeding through genomic



**Figure 1 | Comparison of gene families and coding domains of *T. urartu* with rice, maize, sorghum and *B. distachyon*.** a, Venn diagram illustrating shared and unique gene families (gene numbers in parenthesis) among the five grass species. b, Comparison of protein families with different Pfam domains among the five species. Fourteen Pfam domains that had significant differences ( $P < 0.001$ , Fisher's exact test) in their percentages between *T. urartu* and the other four grass species are shown. The percentage was calculated by dividing the number of each domain by the total gene set in a given species.



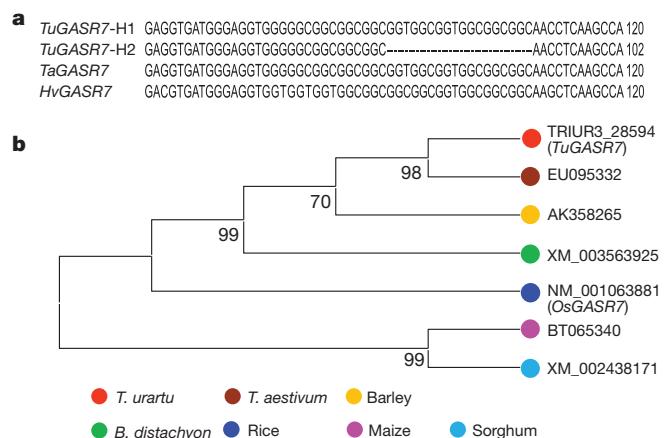
**Figure 2 | Synteny analysis between *T. urartu* and *B. distachyon*, and genome expansion in *T. urartu*.**

**a**, Syntenic relationships between the seven hypothetical chromosomes of *T. urartu* (1A–7A) and the five chromosomes of *B. distachyon* (Bd1–Bd5). The deletion bin maps of bread wheat 1A–7A chromosomes are noted at the bottom. **b**, Comparison of intergenic spaces in one representative syntetic block from *T. urartu* and *B. distachyon*. The 50-kb block from *B. distachyon* chromosome 1 (Bd1) contains five genes (indicated by red arrowheads). The *T. urartu* orthologues (yellow arrowheads) of the five genes are distributed in five separate scaffolds (totalling 1,092 kb). The total intergenic space in the displayed *T. urartu* region is more than 20-fold that in the corresponding Bd1 region.

selection. We identified 739,534 insertion-site-based polymorphism (ISBP) markers and 166,309 simple sequence repeats (SSRs) (Supplementary Information). PCR validation showed that 94.5% of the SSRs and 87% of the ISBP markers gave the expected products, and that

33.61% of the SSRs and 10.19% of the ISBP markers were specific to the A genome. Moreover, 28.7% of the SSR loci were polymorphic in bread wheat (Supplementary Information). To enable the identification of single nucleotide polymorphisms (SNPs), we re-sequenced another *T. urartu* accession (DV2138) and obtained 78.6 Gb of high-quality data. Comparison of the genome data between the two *T. urartu* accessions (G1812 versus DV2138) allowed the discovery of 2,989,540 SNPs, which will be useful for the future development of SNP markers (Supplementary Information).

Previous studies have revealed that more than half of the 60 meta-quantitative trait loci (meta-QTLs)<sup>22,23</sup> related to wheat yield traits are present in the A genome of bread wheat, and three meta-QTLs (MQTL\_5, 6 and 7) are located on chromosome 5A (ref. 22). We therefore searched the *T. urartu* scaffolds using available markers around the three meta-QTLs (Supplementary Information). We found ten scaffolds with a total length of 772,014 bp that were distributed in the 14-centimorgan (cM) region of MQTL\_5; nine scaffolds with a combined length of 783,140 bp were located in the 15-cM region containing MQTL\_6; and six scaffolds with an overall length of 529,604 bp were assigned to the 9-cM region harbouring MQTL\_7. The sequence information of these scaffolds will expedite the development of more polymorphic markers within the three meta-QTL regions and facilitate the identification of their corresponding genes.



**Figure 3 | Association analysis of the *TuGASR7* gene.** **a**, Indel differences between the *TuGASR7* haplotypes H1 and H2 compared with their corresponding regions in the *GASR7* homologues from barley (*Hordeum vulgare*; *HvGASR7*) and *T. aestivum* (*TaGASR7*). **b**, Phylogenetic relationships among the homologous genes of rice *GASR7* from different grass species. TRIUR3\_28594 is the predicted homologue of *GASR7* in *T. urartu*. **c**, Typical grain lengths for the two detected *TuGASR7* haplotypes, H1 (top panel) and H2 (bottom panel). For each haplotype, the seeds from four representative *T. urartu* accessions were photographed. To illustrate the difference in grain length more clearly, ten seeds were randomly selected from each of the compared accessions and aligned closely for the photograph.

Our *T. urartu* draft genome sequence provides new insights into the A genome that is shared by many polyploid wheat species. The large set of gene models (34,879) and abundant genetic markers anchored in sequence scaffolds, together with the emerging genomic resources from bread wheat<sup>8</sup>, promise to accelerate deeper and more systematic genomic and breeding studies of bread wheat that are required to meet the future challenges of food security and sustainable agriculture.

## METHODS SUMMARY

The genome of *T. urartu* accession G1812 was sequenced on the Illumina HiSequation (2000) platform. These data were used to assemble the draft genome sequence with the use of the SOAPdenovo<sup>9</sup> software. RNA-Seq data were generated on the same platform and Roche 454 for genome annotation and transcriptome analysis. Repeat sequences were identified through sequence similarity at the nucleotide and protein levels<sup>24</sup>. Protein-coding genes were predicted by using an *ab initio* approach, sequence similarity search and RNA-Seq data to build reliable gene models<sup>25</sup>. Detailed methodology descriptions are given in Supplementary Information.

Received 17 March 2012; accepted 8 February 2013.

Published online 24 March 2013.

1. Peng, J. H., Sun, D. F. & Nevo, E. Domestication evolution, genetics and genomics in wheat. *Mol. Breed.* **28**, 281–301 (2011).
2. Dvorak, J., di Terlizzi, P., Zhang, H. B. & Resta, P. The evolution of polyploid wheats: identification of the A genome donor species. *Genome* **36**, 21–31 (1993).
3. Dvorak, J. & Zhang, H. B. Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. *Proc. Natl Acad. Sci. USA* **87**, 9640–9644 (1990).
4. McFadden, E. S. & Sears, E. R. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**, 81–89 (1946).
5. Food and Agriculture Organisation of the United Nations. FAOSTAT. <http://faostat.fao.org/site/339/default.aspx> (2011).
6. Paux, E. et al. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* **322**, 101–104 (2008).
7. Hernandez, P. et al. Next-generation sequencing and syntetic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.* **69**, 377–386 (2012).
8. Brenchley, R. et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705–710 (2012).
9. Li, R. et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
10. Bennett, M. D. & Smith, J. B. Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond. B* **274**, 227–274 (1976).
11. Ozkan, H., Tuna, M., Kilian, B., Mori, N. & Ohta, S. Genome size variation in diploid and tetraploid wild wheats. *AoB Plants* **2010**, 1–11 (2010).
12. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
13. Schnable, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
14. Paterson, A. H. et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551–556 (2009).
15. Vogel, J. P. et al. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763–768 (2010).

16. Choulet, F. et al. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* **22**, 1686–1701 (2010).
17. The International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716 (2012).
18. Li, L., Stoeckert, C. J., Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
19. Collier, S. M. & Moffett, P. NB-LRRs work a ‘bait and switch’ on pathogens. *Trends Plant Sci.* **14**, 521–529 (2009).
20. Qi, L. L. et al. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* **168**, 701–712 (2004).
21. Huang, X. et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genet.* **44**, 32–39 (2012).
22. Geges, V. C. et al. A genetic framework for grain size and shape variation in wheat. *Plant Cell* **22**, 1046–1056 (2010).
23. Zhang, L. Y. et al. Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. *J. Integr. Plant Biol.* **52**, 996–1007 (2010).
24. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protocols Bioinform.* **25**, 4.10.1–4.10.14 (2004).
25. Elsik, C. G. et al. Creating a honey bee consensus gene set. *Genome Biol.* **8**, R13 (2007).
26. Ling, H.-Q. et al. Genomic data from *Triticum urartu*—the progenitor of wheat A genome. *GigaScience* <http://dx.doi.org/10.5524/100050> (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank L. Goodman for assistance in editing the manuscript, and M. Bevan and Y.-B. Xue for critical reading of the manuscript. This work was supported by grants from the Ministry of Science and Technology of China (the special fund for the State Key Laboratory of Plant Cell and Chromosome Engineering, 2010DFB33540, 2011CB100304, 2011AA100104 and 2009CB118300).

**Author Contributions** H.-Q.L., A.M.Z., Z.S.L., D.W.W., Ji.W., M.-C.L. and J.D. were responsible for project initiation. Project coordination was by H.-Q.L., A.M.Z., D.W.W., Ju.W. and J.Y.W. The project was managed by H.-Q.L., B.R.Z., Ju.W., Ji.W. and H.M.Y. Data generation and analysis were performed by S.C.Z., D.C.L., H.S., Y.T., D.L., H.J.F., H.L.W., L.L.D., Y.W.L., C.G., Y.C., C.Z., X.S.G., S.S.Z., B.W., K.Y., Q.S.L., W.L.Y., X.Y.L., J.C., X.F.Z., G.B.L., H.J.W., P.Y.X., S.H.Z., X.L., F.M.W., Y.P.Y., X.L.A., Z.Y.D., K.P.Z., J.Y.W., M.J.F., J.B.J., Y.J., J.J.L., Z.B.W. and Y.H.S. The analysis was designed by H.-Q.L., A.M.Z., D.W.W., Q.H.S., D.Z.T., X.J.W., Y.P.T. and X.Q.Z. The paper was written by H.-Q.L., S.C.Z., D.C.L., H.S., A.M.Z. and D.W.W.

**Author Information** This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number AOTI00000000. Sequence assemblies and all short-read data are under project accession numbers SRA030525 (genomic short reads), SRA066084 (resequencing reads), PRJNA182347 (assembly and annotation) and SRA064213 (RNA-Seq). The version described in this paper is the first version, AOTI01000000. Genomic data are also available at the Comprehensive Library for Modern Biotechnology (CLIMB) repository<sup>26</sup>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Ju.W. (wangj@genomics.cn), H.-Q.L. (hqling@genetics.ac.cn), A.M.Z. (amzhang@genetics.ac.cn), D.W.W. (dwwang@genetics.ac.cn) or Z.S.L. (zsli@genetics.ac.cn).

 This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

# Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation

Jizeng Jia<sup>1\*</sup>, Shancen Zhao<sup>2,3\*</sup>, Xiuying Kong<sup>1\*</sup>, Yingrui Li<sup>2\*</sup>, Guangyao Zhao<sup>1\*</sup>, Weiming He<sup>2\*</sup>, Rudi Appels<sup>4\*</sup>, Matthias Pfeifer<sup>5</sup>, Yong Tao<sup>2</sup>, Xueyong Zhang<sup>1</sup>, Ruilian Jing<sup>1</sup>, Chi Zhang<sup>2</sup>, Youzhi Ma<sup>1</sup>, Lifeng Gao<sup>1</sup>, Chuan Gao<sup>2</sup>, Manuel Spannagl<sup>5</sup>, Klaus F. X. Mayer<sup>5</sup>, Dong Li<sup>2</sup>, Shengkai Pan<sup>2</sup>, Fengya Zheng<sup>2,3</sup>, Qun Hu<sup>6</sup>, Xianchun Xia<sup>1</sup>, Jianwen Li<sup>2</sup>, Qinsi Liang<sup>2</sup>, Jie Chen<sup>2</sup>, Thomas Wicker<sup>7</sup>, Caiyun Gou<sup>2</sup>, Hanhui Kuang<sup>6</sup>, Genyun He<sup>2</sup>, Yadan Luo<sup>2</sup>, Beat Keller<sup>7</sup>, Qiuju Xia<sup>2</sup>, Peng Lu<sup>2</sup>, Junyi Wang<sup>2</sup>, Hongfeng Zou<sup>2</sup>, Rongzhi Zhang<sup>1</sup>, Junyang Xu<sup>2</sup>, Jinlong Gao<sup>2</sup>, Christopher Middleton<sup>7</sup>, Zhiwu Quan<sup>2</sup>, Guangming Liu<sup>8</sup>, Jian Wang<sup>2</sup>, International Wheat Genome Sequencing Consortium†, Huanming Yang<sup>2</sup>, Xu Liu<sup>1</sup>, Zhonghu He<sup>1,9</sup>, Long Mao<sup>1</sup> & Jun Wang<sup>2,10,11</sup>

**About 8,000 years ago in the Fertile Crescent, a spontaneous hybridization of the wild diploid grass *Aegilops tauschii* ( $2n = 14$ ; DD) with the cultivated tetraploid wheat *Triticum turgidum* ( $2n = 4x = 28$ ; AABB) resulted in hexaploid wheat (*T. aestivum*;  $2n = 6x = 42$ ; AABBDD)<sup>1,2</sup>. Wheat has since become a primary staple crop worldwide as a result of its enhanced adaptability to a wide range of climates and improved grain quality for the production of baker's flour<sup>2</sup>. Here we describe sequencing the *Ae. tauschii* genome and obtaining a roughly 90-fold depth of short reads from libraries with various insert sizes, to gain a better understanding of this genetically complex plant. The assembled scaffolds represented 83.4% of the genome, of which 65.9% comprised transposable elements. We generated comprehensive RNA-Seq data and used it to identify 43,150 protein-coding genes, of which 30,697 (71.1%) were uniquely anchored to chromosomes with an integrated high-density genetic map. Whole-genome analysis revealed gene family expansion in *Ae. tauschii* of agronomically relevant gene families that were associated with disease resistance, abiotic stress tolerance and grain quality. This draft genome sequence provides insight into the environmental adaptation of bread wheat and can aid in defining the large and complicated genomes of wheat species.**

We selected *Ae. tauschii* accession AL8/78 for genome sequencing because it has been extensively characterized genetically (Supplementary Information). Using a whole genome shotgun strategy, we generated 398 Gb of high-quality reads from 45 libraries with insert sizes ranging from 200 bp to 20 kb (Supplementary Information). A hierarchical, iterative assembly of short reads employing the parallelized sequence assembler SOAPdenovo<sup>3</sup> achieved contigs with an N50 length (minimum length of contigs representing 50% of the assembly) of 4,512 bp (Table 1). Paired-end information combined with an additional 18.4 Gb of Roche/454 long-read sequences was used sequentially to generate 4.23-Gb scaffolds (83.4% were non-gapped contiguous sequences) with an N50 length of 57.6 kb (Supplementary Information). The assembly represented 97% of the 4.36-Gb genome as estimated by K-mer analysis (Supplementary Information). We also obtained 13,185 *Ae. tauschii* expressed sequence tag (EST) sequences using Sanger sequencing, of which 11,998 (91%) could be mapped to the scaffolds with more than 90% coverage (Supplementary Information).

To aid in gene identification, we performed RNA-Seq (53.2 Gb for a 117-Mb transcriptome assembly) on 23 libraries representing eight tissues including pistil, root, seed, spike, stamen, stem, leaf and sheath

(Supplementary Information). Using both evidence-based and *de novo* gene predictions, we identified 34,498 high-confidence protein-coding loci. FGENESH<sup>4</sup> and GeneID models were supported by a 60% overlap with either our ESTs and RNA-Seq reads, or with homologous proteins. More than 76% of the gene models had a significant match ( $E$  value  $\leq 10^{-5}$ ; alignment length  $\geq 60\%$ ) in the GenBank non-redundant database. An additional 8,652 loci were predicted as low-confidence genes as a result of incomplete gene structure or limited expression data support (Supplementary Information). We also predicted a total of 2,505 transfer RNA, 358 ribosomal RNA, 35 small nuclear RNA and 78 small nucleolar RNA genes (Supplementary Information).

We found that more than 65.9% of the *Ae. tauschii* genome was composed of different transposable element (TE) families (Supplementary Information). About  $5 \times 10^6$  Illumina reads of *Ae. tauschii* were mapped to hexaploid wheat repetitive sequences and we found that a comparable percentage of reads (more than 62.3%) could be classified as part of a TE sequence (Supplementary Fig. 6). This estimate is similar to that derived from a previous survey of Roche/454 sequences<sup>5</sup>. There were 410 different TE families, of which the 20 most abundant contributed more than 50% of the *Ae. tauschii* genome (Supplementary Table 9). A single peak of increased insertion activity was estimated to occur about 3–4 Myr ago by measuring the similarity of the assembled LTR retrotransposons (Supplementary Information), suggesting that the expansion of the *Ae. tauschii* genome was relatively recent and coincided with the abrupt climate change during the Pliocene Epoch<sup>6</sup>.

We constructed a high-density genetic map using an  $F_2$  population of 490 individuals derived from a cross between the *Ae. tauschii* accessions Y2280 and AL8/78. The map, whose total length was 1059.8 centimorgans (cM), consisted of 151,083 single nucleotide polymorphism (SNP) markers developed by restriction-site-associated DNA (RAD) tag sequencing technology (Supplementary Fig. 13). Together with bin-mapped wheat ESTs<sup>7</sup>, SNPs and tags<sup>8</sup>, the genetic map was used to align 30,303 scaffolds (1.72 Gb; 30,697 genes) to chromosomes (Supplementary Information). The *Ae. tauschii* genes and scaffolds were also anchored to barley<sup>9</sup> and *Brachypodium* chromosome maps<sup>10</sup> (Fig. 1 and Supplementary Fig. 17). Calculation of  $K_a/K_s$  ratios (the ratio of non-synonymous substitutions to synonymous substitutions) for pairs of conserved orthologous genes showed that the average values between *Ae. tauschii* and barley (20,892 genes), *Brachypodium* (17,231 genes), rice (16,370 genes) and sorghum (18,623 genes) were 0.2214, 0.1888, 0.1736 and 0.1726, respectively.

<sup>1</sup>National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China. <sup>2</sup>BGI-Shenzhen, Shenzhen 518083, China. <sup>3</sup>State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. <sup>4</sup>Centre for Comparative Genomics, Murdoch University, Perth, WA 6150, Australia. <sup>5</sup>MIPS/Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, D-85764 Neuherberg, Germany. <sup>6</sup>Department of Vegetable Crops, College of Horticulture and Forestry, Huazhong Agricultural University, Wuhan 430070, China. <sup>7</sup>Institute of Plant Biology, University of Zurich, CH-8008 Zurich, Switzerland. <sup>8</sup>National Supercomputer Center in Tianjin, Tianjin 300457, China. <sup>9</sup>International Maize and Wheat Improvement Center (CIMMYT), Texcoco CP 56130, Mexico. <sup>10</sup>Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark. <sup>11</sup>King Abdulaziz University, Jeddah 21589, Saudi Arabia.

\*These authors contributed equally to this work.

†A list of participants and their affiliations appears in the Supplementary Information.

**Table 1 | Overall statistics of sequencing and genome assembly**

| Assembly process | Library insert size (bp) | Read length (bp)                     | Effective data (Gb) | N50 (bp) | N50 number | Total length (Mb) | Gaps (Mb) |
|------------------|--------------------------|--------------------------------------|---------------------|----------|------------|-------------------|-----------|
| Contig assembly  | 167–764                  | 44, 75, 100, 150                     | 270                 | 4,521    | 179,145    | 3,528             | –         |
| Scaffolding      | 2,000–20,000             | 44, 49, 90                           | 128                 | 58,011   | 19,405     | 4,244             | 1,122     |
| Gap closure      | 167–764                  | 44, 75, 100, 150<br>114–263<br>~600* | 270<br>65<br>18     | 57,585   | 19,455     | 4,229             | 701       |

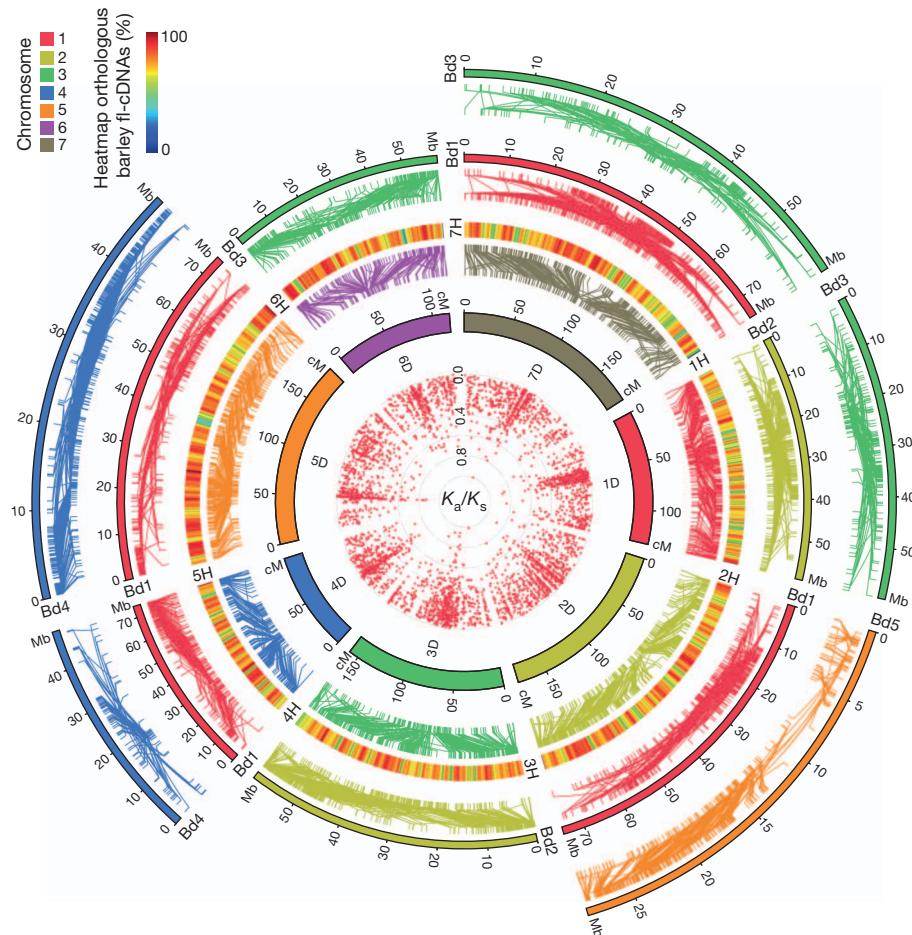
\* Reads from 454 sequencing platform.

which indicated that most gene lineages evolved under purifying selection in *Ae. tauschii*. A total of 628 genes exhibited  $K_a/K_s$  ratios of more than 0.8 when compared with the other four species, indicating potential positive selection (innermost circle of Fig. 1). These genes were assigned to a wide range of molecular functions by using Gene Ontology (GO) analyses (Supplementary Table 14).

*Ae. tauschii* proteins were clustered with those of *Brachypodium*, rice, sorghum and barley (full-length complementary DNAs), and formed 23,202 orthologous groups (at least two members; Supplementary Information). In total, we identified 11,289 (barley/*Ae. tauschii*) and 14,675 (*Brachypodium/Ae. tauschii*) orthologous gene pairs. We found that 8,443 gene groups contained sequences from all five grass genomes, and 234 were specific to Pooideae (*Ae. tauschii*, *Brachypodium* and barley) and 587 were specific to Triticeae (*Ae. tauschii* and barley) (Fig. 2a). Enrichment analyses of both Pfam domains and GO terms showed that genes encoding NBS-LRR (nucleotide-binding-site

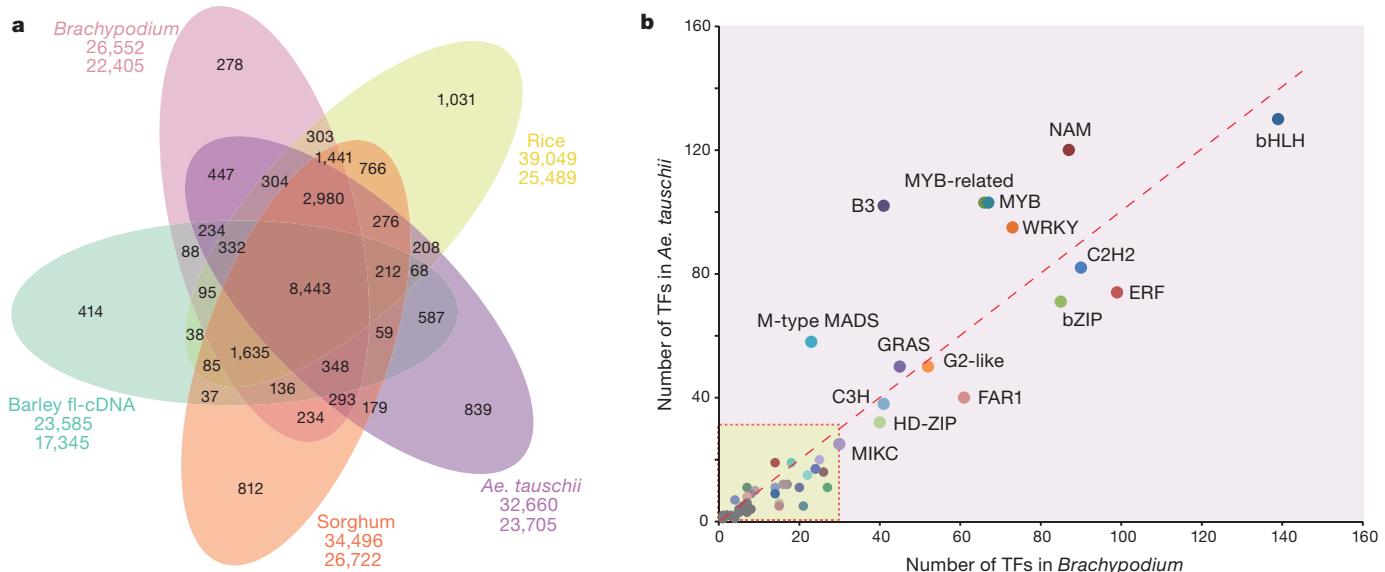
leucine-rich repeat) proteins were over-represented in *Ae. tauschii* relative to *Brachypodium* and rice<sup>11,12</sup> (Supplementary Information). These observations are consistent with those reported in a recent study<sup>13</sup>. A total of 1,219 *Ae. tauschii* genes were similar to NBS-LRR genes (R gene analogues (RGAs))<sup>11,14</sup> (Supplementary Information). This number is double that in rice (623) and sixfold that in maize (216)<sup>12</sup>, indicating that the RGA family has substantially expanded in *Ae. tauschii*. We mapped 878 RGAs (72%) to specific positions across wheat chromosomes by using molecular marker–genome sequence alignment, which provides a large number of potential disease resistance loci for further investigation.

We found more genes for the cytochrome P450 family in *Ae. tauschii* (485) than in sorghum (365), rice (333), *Brachypodium* (262) or maize (261). This family of genes is important for abiotic stress response, especially in biosynthetic and detoxification pathways<sup>15</sup>. Using 178 manually curated cold-acclimation-related genes such



**Figure 1 | Comparative analysis of *Ae. tauschii* ordered scaffolds versus barley and *Brachypodium*.** The inner circle represents the seven *Ae. tauschii* chromosomes scaled according to the genetic map incorporating genome scaffolds. Red points show the  $K_a/K_s$  ratios between anchored *Ae. tauschii* genes and their putative orthologues in *Brachypodium*. Moving outwards, the second circle compares *Ae. tauschii* against the seven barley chromosomes<sup>9</sup>. The

heatmaps show the density distribution of barley cDNA loci that are aligned with *Ae. tauschii* genes. The outer two circles illustrate *Brachypodium* chromosomes according to conserved synteny with *Ae. tauschii*. The coloured lines below each chromosome identify putative orthologous gene pairs between *Ae. tauschii* genes, barley genes and *Brachypodium* genes.



**Figure 2 | Ae. tauschii gene families and transcription factors.**

**a**, Distribution of orthologous gene families in *Ae. tauschii*, *Brachypodium*, sorghum, rice and barley. The number of gene families is represented in each intersection of the Venn diagram. The first number under the species name indicates the total number of genes annotated for a particular species, and the

second indicates the number of genes in groups for that organism. The difference between the two accounts for singleton genes that were not present in any cluster. **b**, The composition of transcription factors (TFs) in *Ae. tauschii* and *Brachypodium* composed of more than 30 members.

as the CCAAT-binding factor (CBF) transcription factors<sup>16</sup>, late-embryogenesis-abundant proteins (LEA) and osmoprotectant biosynthesis proteins (Supplementary Information) as queries, we identified 216 cold-related genes in the *Ae. tauschii* genome, in contrast to 164 genes in *Brachypodium*, 132 in rice, 159 in sorghum and 148 in maize. Some of these genes were specific to *Ae. tauschii* or to Pooideae, including those encoding ice-recrystallization inhibition protein 1 precursor, DREB2 transcription factor  $\alpha$  isoform and cold-responsive LEA/RAB-related COR protein. Expression analysis of RNA-Seq data showed that most of these *Ae. tauschii*-specific and Pooideae-specific genes were constitutively expressed in *Ae. tauschii* (Supplementary Fig. 23). In addition, 1,489 transcription factors (TFs) in 56 families were identified by using Pfam DNA-binding domains (Supplementary Information). *Ae. tauschii* had an excess of such TFs as MYB-related genes (103, in contrast to 66 in *Brachypodium* and 95 in maize), and these are also thought to be involved in various stress responses<sup>17</sup>. The M-type MADS-box genes (58, in contrast with 23 in *Brachypodium* and 34 in maize) are involved in regulation in plant reproduction<sup>18</sup> (Fig. 2b and Supplementary Table 18). ARACNe<sup>19</sup> co-expression analysis using RNA-Seq data predicted an expression network of 1,283 interactions (Supplementary Fig. 25), in which 13 TFs were associated with the expression of drought tolerance genes<sup>20</sup> (Supplementary Table 20).

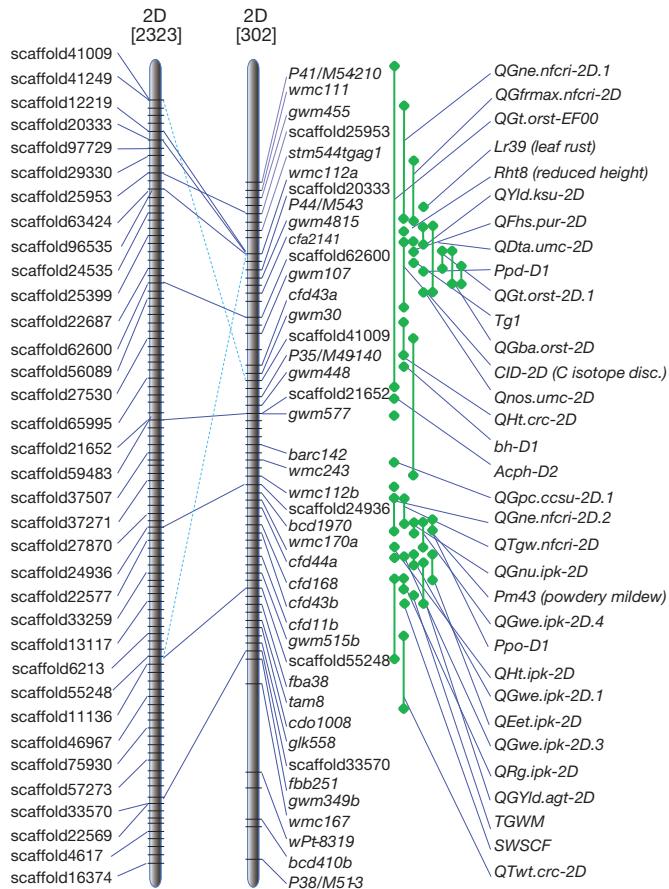
We predicted a total of 159 (133 families) previously undescribed microRNAs (Supplementary Information), and identified segmental and tandem duplications in 42 members of the miR2118 family that were organized into two groups on 15 scaffolds (Supplementary Fig. 26). The miR399 family, which is involved in the regulation of inorganic phosphate homeostasis in rice<sup>21</sup>, was expanded (20 members in *Ae. tauschii*, compared with 11 in rice and 10 in maize), and may contribute to the ability of *Ae. tauschii* to grow in low-nutrient soils. The expansion of the miR2275 family (eight members in *Ae. tauschii*, compared with two in rice and four in maize) may contribute to the enhanced disease resistance of *Ae. tauschii* because phased short interfering RNAs initiated by miR2275 have been implicated in these activities<sup>22</sup>.

The *Ae. tauschii* genome served as the source for many grain quality genes in hexaploid wheat, creating a step improvement in the formation of the elastic dough essential for bread making<sup>2</sup>. Grain quality genes include high-molecular-weight glutenin subunits (HMW-GS),

low-molecular-weight glutenin subunits (LMW-GS)<sup>23</sup>, grain texture proteins (GSP; puroindolines)<sup>24</sup> and storage protein activator (SPA)<sup>25</sup>. We identified two HMW-GS genes, five LMW-GS genes, one *Pina* gene, two *Pinb* genes, one *GSP* gene and one *SPA* gene in the *Ae. tauschii* genome sequence (Supplementary Information). As has been shown for the Hardness (*Ha*) locus<sup>24</sup>, the *GSP*, *Pina* and *Pinb* genes were also organized in a cluster. RNA-Seq analysis showed that these grain quality genes were expressed predominantly in seeds (Supplementary Fig. 29).

The anchoring of more than 40% of the scaffold sequences to four genetic maps and to syntenic regions of other sequenced grass species provided a structural framework for integrating multiple maps by using shared markers (Fig. 1 and Supplementary Information). The co-localization of genes in scaffolds and genetically mapped quantitative trait loci (QTLs) will directly support map-based gene cloning. On chromosome 2D, for example, the locations of 33 QTLs or genes were integrated with scaffold information (<http://ccg.murdoch.edu.au/cmap/ccg-live/>) (Fig. 3 and Supplementary Information). Alignment of the *Ae. tauschii* genetic map with the wheat 2D consensus genetic map was unambiguous, with the exception of some single crossovers that were probably due to repetitive elements (dotted lines in Fig. 3). The genome sequence also provided the basis for the identification of more than 860,126 simple sequence repeats (SSRs), with trimers (37.7%) and tetramers (27.5%) as the most abundant SSR types (Supplementary Information). Together with the 711,907 SNPs identified by resequencing a roughly fivefold coverage of a second accession, Y2280 (Supplementary Information), the genomic resources reported here will promote map-based gene cloning and marker-assisted selection in wheat.

With its high base accuracy and nearly complete set of gene sequences, the *Ae. tauschii* draft genome sequence provides an essential reference for studying D genome diversity by re-sequencing additional accessions. Over the past half century, the introduction of new D genome diversity into synthetic wheat has been a major effort to expand bread wheat genetic diversity and to create environmentally resilient lines<sup>26,27</sup>. The *Ae. tauschii* genome sequence should aid in identifying new elite alleles for agriculturally important traits to alleviate the worsening plight of global climate and environment changes<sup>27</sup>.



**Figure 3 | An integrated genetic map of *Ae. tauschii* chromosome 2D.** The *Ae. tauschii* genetic map was integrated with markers, scaffolds and mapped QTLs to assist in marker development and map-based cloning. Left: the *Ae. tauschii* molecular map used for synteny alignment in Fig. 1 was aligned to chromosome 2D (November 2011 consensus map, CMap; <http://ccg.murdoch.edu.au/cmap/ccg-live>) where sequence information was available. The original marker at a location is retained in CMap as a synonym. Right: within CMap, details for QTL locations are provided at a greater magnification to show all the markers in the regions of interest. The dotted lines indicate an ambiguous relationship that is most probably due to repetitive sequences.

## METHODS SUMMARY

We selected *Ae. tauschii* ( $2n = 14$ ) accession AL8/78 for sequencing. Plants were grown at  $25^{\circ}\text{C}$  in a darkened chamber for two weeks; DNA was extracted from leaf tissue and purified with a standard phenol/chloroform extraction protocol. Sequencing libraries were constructed and sequenced on Illumina next-generation sequencing platforms (GAII and HiSequation (2000)). High-quality reads were assembled with SOAPdenovo<sup>3</sup>. Repeat sequences were identified by combining *de novo* approaches and sequence similarity at the nucleotide and protein levels. Gene models were predicted by combining homology-based, *de novo* and RNA-Seq-based methods. RNA-Seq reads were assembled with CAP3 (ref. 28) and CD-Hit<sup>29</sup> and were mapped to the draft genome with Tophat<sup>30</sup>. See Supplementary Information for details and additional analyses.

Received 28 March 2012; accepted 20 February 2013.

Published online 24 March; corrected online 3 April 2013 (see full-text HTML version for details).

1. Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. & Martin, W. Genetics and geography of wild cereal domestication in the near east. *Nature Rev. Genet.* **3**, 429–441 (2002).
2. Dubcovsky, J. & Dvorak, J. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**, 1862–1866 (2007).
3. Li, R. et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
4. Salamov, A. A. & Solovyev, V. V. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).

5. Wicker, T. et al. A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant J.* **59**, 712–722 (2009).
6. Williams, M. et al. Pliocene climate and seasonality in North Atlantic shelf seas. *Phil. Trans. R. Soc. A* **367**, 85–108 (2009).
7. Qi, L. L. et al. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* **168**, 701–712 (2004).
8. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**, e32253 (2012).
9. Mayer, K. F. et al. Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* **23**, 1249–1263 (2011).
10. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763–768 (2010).
11. McHale, L., Tan, X., Koehl, P. & Michelmore, R. W. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **7**, 212 (2006).
12. Luo, S. et al. Dynamic nucleotide-binding-site and leucine-rich-repeat-encoding genes in the grass family. *Plant Physiol.* **159**, 197–210 (2012).
13. Brenchley, R. et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705–710 (2012).
14. Yue, J. X., Meyers, B. C., Chen, J. Q., Tian, D. & Yang, S. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol.* **193**, 1049–1063 (2012).
15. Schuler, M. A. & Werck-Reichhart, D. Functional genomics of P450s. *Annu. Rev. Plant Biol.* **54**, 629–667 (2003).
16. Thomashow, M. F. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiol.* **154**, 571–577 (2010).
17. Lata, C., Yadav, A. & Prasad, M. In *Abiotic Stress Response in Plants—Physiological, Biochemical and Genetic Perspectives* (Shanker, A. & Venkateswarlu, B., eds) 269–296 (InTech, 2011).
18. Masiero, S., Colombo, L., Grini, P. E., Schnittger, A. & Kater, M. M. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* **23**, 865–872 (2011).
19. Margolin, A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7** (Suppl. 1), S7 (2006).
20. Mao, X. et al. Transgenic expression of *TaMYB2A* confers enhanced tolerance to multiple abiotic stresses in *Arabidopsis*. *Funct. Integr. Genomics* **11**, 445–465 (2011).
21. Hu, B. & Chu, C. Phosphate starvation signaling in rice. *Plant Signal. Behav.* **6**, 927–929 (2011).
22. Shivaprasad, P. V. et al. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* **24**, 859–874 (2012).
23. Gupta, R. B., Singh, N. K. & Shepherd, K. W. The cumulative effect of allelic variation in LMW and HMW glutenin subunits on dough properties in the progeny of two bread wheats. *Theor. Appl. Genet.* **77**, 57–64 (1989).
24. Chantret, N. et al. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* **17**, 1033–1045 (2005).
25. Ravel, C. et al. Nucleotide polymorphism in the wheat transcriptional activator *Spa* influences its pattern of expression and has pleiotropic effects on grain protein composition, dough viscoelasticity, and grain hardness. *Plant Physiol.* **151**, 2133–2144 (2009).
26. Talbert, L. E., Smith, L. Y. & Blake, N. K. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome* **41**, 402–407 (1998).
27. Trethowan, R. M. & Mujeeb-Kazi, A. Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. *Crop Sci.* **48**, 1255–1265 (2008).
28. Huang, X. & Madan, A. CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
29. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
30. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. M. Wan for support and encouragement; J. Dvorak and M. C. Luo for the AL8/78 line; C. Y. Jin, X. Y. Li, L. C. Zhang, L. Pan and J. C. Zhang for material preparation; Y. H. Lv for providing helpful palaeogeological information; D. M. Appels for producing the CMap database of molecular genetic maps; K. Edwards for providing the details of the SNP-based map for Avalon × Cadenza; L. Goodman for assistance in editing the manuscript; and M. W. Bevan, Y. B. Xu and C. Zou for critical readings of the manuscript. This work was supported by grants from the National 863 Project (2012AA10A0308 and 2011AA100104), the International S&T Cooperation Program of China (2008DFB30080), the National Natural Science Foundation of China (31171548 and 31071415), the National Basic Research Program of China (2010CB125900), the Core Research Budget of the Non-profit Governmental Research (201013) and the National Program on R&D of Transgenic Plants (2011ZX08009-001 and 2011ZX08002-002).

**Author Contributions** J.Z.J., Ji.W., Xu.L., H.M.Y., Z.H.H., Lo.M., Ju.W., X.Y.K., X.Y.Z., R.L.J. and Y.Z.M. initiated the project and designed the study. X.Y.K., G.Y.Z., R.A., L.F.G., Qu.H., H.H.K., B.K., X.C.X., R.Z.Z., G.M.L. and C.M. performed the research. S.C.Z., Y.R.L., W.M.H., M.P., C.Z., D.L., C.G., M.S., K.M., Y.T., F.Y.Z., S.K.P., J.W.L., Q.S.L., Ji.C., C.Y.G., G.Y.H., Y.D.L., P.L., J.Y.W., J.Y.X. and J.L.G. generated and analysed the data. Q.J.X., H.F.Z. and Z.W.Q. developed the high-density genetic map. J.Z.J., S.C.Z., Lo.M., R.A., G.Y.Z., X.Y.K. and T.W. wrote the paper.

**Author Information** The genome sequence and the annotation are available from the National Centre for Biotechnology Information (NCBI) as BioProject ID PRJNA182898. This Whole Genome Shotgun project is deposited at DDBJ/EMBL/GenBank under accession number AOC0000000000. The version described in this paper is the first version, AOC001000000. The Illumina sequencing reads are available in the Sequence Read Archive under accession number SRA030526, RNA-Seq sequences under SRA062662, and resequencing short reads under SRA063175. Genomic data are also available at the Comprehensive Library for Modern Biotechnology (CLiMB) repository under doi:10.5524/100054. Reprints and permissions information is

available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Ju.W. (wangj@genomics.cn), J.Z.J. (jzjia@mail.caas.net.cn), Lo.M. (maolong@caas.net.cn), Z.H.H. (zhhe@public3.bta.net.cn) and Xu L. (liuxu01@caas.cn).

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

# The emergence of functional microcircuits in visual cortex

Ho Ko<sup>1\*</sup>, Lee Cossell<sup>1\*</sup>, Chiara Baragli<sup>1</sup>, Jan Antolik<sup>1†</sup>, Claudia Clopath<sup>2</sup>, Sonja B. Hofer<sup>1\*</sup> & Thomas D. Mrsic-Flogel<sup>1\*</sup>

**Sensory processing occurs in neocortical microcircuits in which synaptic connectivity is highly structured<sup>1–4</sup> and excitatory neurons form subnetworks that process related sensory information<sup>5,6</sup>. However, the developmental mechanisms underlying the formation of functionally organized connectivity in cortical microcircuits remain unknown. Here we directly relate patterns of excitatory synaptic connectivity to visual response properties of neighbouring layer 2/3 pyramidal neurons in mouse visual cortex at different postnatal ages, using two-photon calcium imaging *in vivo* and multiple whole-cell recordings *in vitro*. Although neural responses were already highly selective for visual stimuli at eye opening, neurons responding to similar visual features were not yet preferentially connected, indicating that the emergence of feature selectivity does not depend on the precise arrangement of local synaptic connections. After eye opening, local connectivity reorganized extensively: more connections formed selectively between neurons with similar visual responses and connections were eliminated between visually unresponsive neurons, but the overall connectivity rate did not change. We propose a sequential model of cortical microcircuit development based on activity-dependent mechanisms of plasticity whereby neurons first acquire feature preference by selecting feedforward inputs before the onset of sensory experience—a process that may be facilitated by early electrical coupling between neuronal subsets<sup>7–9</sup>—and then patterned input drives the formation of functional subnetworks through a redistribution of recurrent synaptic connections.**

Intrinsic and experiential factors guide the patterning of neural pathways and the establishment of sensory response properties during postnatal development<sup>10–12</sup>. During this time, neural circuit refinement is thought to depend on the elimination of initially exuberant projections, selective formation of new connections or both<sup>13,14</sup>. However, the mechanisms governing the emergence of structured connectivity in local cortical microcircuits, where dendrites and axons overlap extensively, remain uncertain. Moreover, it is not clear whether the organization of synaptic connections between nearby neurons is established early and inherently linked to the formation of receptive fields (RFs) before the onset of sensory experience or whether the mature patterns of recurrent connectivity appear only after the formation of RFs as a result of correlated activity induced by feedforward drive from the sensory periphery. Here we investigate these questions in networks of layer 2/3 (L2/3) pyramidal cells in mouse primary visual cortex (V1)—where neighbouring neurons exhibit a diversity of preference for visual features<sup>5,15–17</sup>—by determining how local synaptic connectivity relates to visual response properties during development.

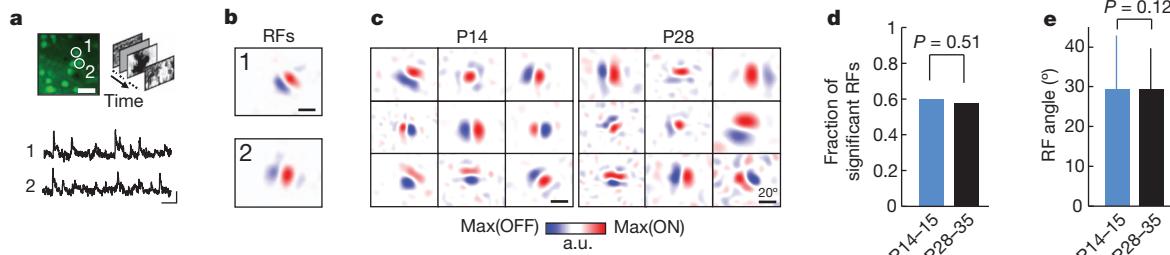
Stimulus-selective responses in V1 are observed at eye opening<sup>18–22</sup>, but the extent to which detailed RFs are established by this time remains unclear. To characterize the spatial RF structure of L2/3 neurons in V1 at eye opening (postnatal days (P) 14–15) and in more mature mice (P28–35), we used *in vivo* two-photon calcium imaging<sup>23</sup>

in monocular V1 to obtain spatial RFs by reverse correlation<sup>16,17,24</sup> of static natural images and spiking responses inferred from calcium signals<sup>25</sup> (Methods and Fig. 1a, b), and fitted a two-dimensional Gabor function to the RFs (Supplementary Fig. 1). At eye opening, L2/3 neurons exhibited a diversity of RF spatial structures that resembled those of mature V1 neurons (Fig. 1c and Supplementary Fig. 1a). The proportion of neurons with significant linear RFs was similar between the two age groups (P14–15, 60%, 191 of 317; P28–35, 58%, 201 of 348;  $P = 0.51$ ,  $\chi^2$  test; Fig. 1d), as was the angle of visual space subtended by RFs (mean visual angle along the long RF axis  $\pm$  s.d.: P14–15,  $29.3 \pm 13.6^\circ$ ; P28–35,  $29.4 \pm 10.3^\circ$ ;  $P = 0.12$ , rank-sum test; Fig. 1e; see also Supplementary Fig. 1b, c). The similarity of RF structures was shown by the overlapping distributions of standard RF measures of  $n_x$  and  $n_y$ , which respectively express the width and length of the fitted Gabor function in units of the underlying grating period (median  $n_x$ : P14–15, 0.31; P28–35, 0.32;  $P = 0.14$ ; median  $n_y$ : P14–15, 0.20; P28–35, 0.20;  $P = 0.41$ ; rank-sum test; Methods and Supplementary Fig. 1d). For comparison, the orientation tuning of neurons responsive to drifting gratings was slightly but significantly broader at eye opening than in more mature V1 (median orientation selectivity index: P13–15, 0.62; P22–26, 0.68;  $P = 2.39 \times 10^{-34}$ , rank-sum test; Supplementary Fig. 2a, b). Neuronal responses at eye opening, however, exhibited higher variability to repeated presentation of the preferred grating stimulus (coefficient of variation (CV)) than in more mature animals (mean CV  $\pm$  s.d.: P13–15,  $0.93 \pm 0.36$ ; P22–26,  $0.71 \pm 0.30$ ;  $P = 1.17 \times 10^{-304}$ , rank-sum test; Supplementary Fig. 2c), indicating that responses in immature animals were less reliable despite the presence of clearly defined RFs and orientation selectivity.

The similarity of spatial RF properties in immature and older V1 suggests that the organization of feedforward connections<sup>26</sup> was largely established by the time of eye opening. We next tested whether the synaptic connectivity of neurons in the local cortical network is functionally specific when vision begins. We combined *in vivo* two-photon calcium imaging in V1 and subsequent multiple whole-cell recordings in slices of the same tissue<sup>5</sup> (Fig. 2). We first imaged calcium signals at consecutive depths within L2/3 to characterize the responses to natural movies and drifting gratings (Methods) of all neurons within a volume of approximately  $285 \mu\text{m} \times 285 \mu\text{m} \times 40\text{--}120 \mu\text{m}$ . We then carried out simultaneous whole-cell recordings *in vitro* from two to six neighbouring L2/3 pyramidal neurons separated by less than  $50 \mu\text{m}$  (mean distance  $\pm$  s.d.: P13–15,  $24 \pm 9 \mu\text{m}$ ; P22–26,  $25 \pm 10 \mu\text{m}$ ; Fig. 2a, b). We recorded from 143 and 140 neurons in total in the slices at P13–15 and P22–26, respectively, which were identified in the *in vivo* image stack by image registration based on affine transformation<sup>5</sup> (Fig. 2a). Synaptic connectivity was assessed by evoking action potentials in each neuron sequentially while searching for the presence of excitatory postsynaptic potentials in the other neurons (Fig. 2b). This approach allowed us to relate the probability of finding connections between pairs of L2/3 neurons to the correlation of their average responses to

<sup>1</sup>Department of Neuroscience, Physiology and Pharmacology, University College London, 21 University Street, London WC1E 6DE, UK. <sup>2</sup>Center for Theoretical Neuroscience, Columbia University, 1051 Riverside Drive, Unit 87 Kolb Research Annex, New York, New York 10032, USA. <sup>†</sup>Present address: Unité de Neuroscience Information et Complexité, UPR 3293 CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France.

\*These authors contributed equally to this work.



**Figure 1 | Responses of L2/3 pyramidal cells in mouse visual cortex are highly feature selective at eye opening.** **a**, Example of OGB-labelled region at P14 (left; scale bar, 30  $\mu$ m) with calcium transients of two cells obtained with two-photon microscopy (bottom; scale bars, 20 s, 10%  $\Delta F/F$ ) in response to natural image sequences. **b**, Linear receptive fields (RFs) of the neurons in

**a** obtained by regularized reverse correlation (Methods). Scale bars, 20°. **c**, RFs of neurons from two mice at different ages. a.u., arbitrary units; red indicates ON subfield; blue indicates OFF subfield. **d**, **e**, Fraction of neurons with significant RFs (**d**,  $\chi^2$  test) and RF size (**e**, rank-sum test) at eye opening and in more mature V1. Error bars, s.d.; P14–15,  $n = 4$  mice; P28–35,  $n = 5$  mice.

natural movies (signal correlation; Fig. 2c, d, f, h) and to the differences in their preferred orientation (Fig. 2g, i). We used natural movie signal correlation for comparison of response similarity not only because it was a good predictor of the similarity of the neurons' linear RFs (Supplementary Fig. 3), but because it also probably captures the similarity of feature selectivity in neurons with nonlinear RFs, which could not be estimated by reverse correlation.

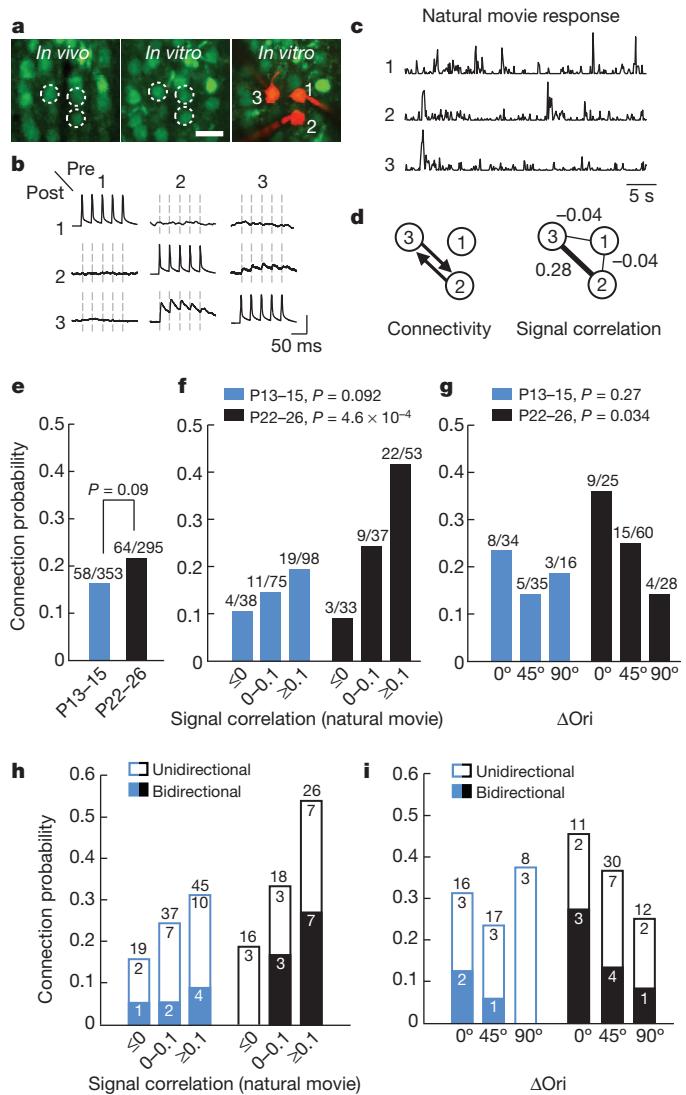
The overall rate of connectivity was not significantly different between the two age groups (P13–15, 16.4%, 58 of 353 connections tested; P22–26, 21.7%, 64 of 295 connections tested;  $P = 0.09$ ,  $\chi^2$  test; Fig. 2e). Among the neurons recorded *in vitro*, 73% (104 of 143) and 56% (79 of 140) exhibited significant responses to the natural movie *in vivo* (Methods). As we reported previously<sup>5</sup>, among responsive L2/3 pyramidal cells the connection probability increased steeply with increasing signal correlation at P22–26 ( $P = 4.6 \times 10^{-4}$ , Cochran–Armitage test; Fig. 2f). This was not the case at P13–15 ( $P = 0.092$ ), when a much weaker trend was observed. Specifically, there were twice as many connections between highly correlated neuronal pairs (signal correlation,  $\geq 0.1$ ) in older V1 than at eye opening (P13–15, 19.4%, 19 of 98 tested; P22–26, 41.5%, 22 of 53 tested;  $P = 0.0035$ ,  $\chi^2$  test; Fig. 2f). Therefore, the functional selectivity of synaptic connections increased in the period after eye opening, as more connections formed selectively between neurons responding to similar stimulus features.

We further examined the refinement of connection specificity by relating the connection rate between reliably responsive, orientation-tuned neurons (P13–15, 43.4%, 62 of 143 neurons; P22–26, 57.9%, 81 of 140; Fig. 2g and Methods) to the difference in their preferred orientation. A significant decreasing relationship between connection probability and the difference in preferred orientation was present in more mature V1 but not at eye opening (P13–15,  $P = 0.27$ ; P22–26,  $P = 0.034$ ; Cochran–Armitage test; Fig. 2g). Together, these results indicate that at eye-opening the organization of synaptic connections between nearby L2/3 pyramidal neurons exhibits only weak functional specificity. After the onset of visual experience, connectivity increases specifically between neurons coding for similar visual features.

Previous studies suggest that bidirectional recurrent connections are overexpressed in some cortical networks<sup>2,4</sup> and that they are most frequent between L2/3 pyramidal cell pairs with similar visual responses in mature V1<sup>5</sup>. We examined whether a similar organization of bidirectional motifs is already present at eye-opening. In contrast to mature mice, visually naive mice did not exhibit a larger proportion of bidirectionally connected pairs between neurons with highly correlated responses to the natural movie (P13–15,  $P = 0.27$ ; P22–26,  $P = 0.01$ ; Cochran–Armitage test; Fig. 2h). Similar trends were found between neurons preferring similar orientations (P13–15,  $P = 0.13$ ; P22–26,  $P = 0.11$ ; Cochran–Armitage test; Fig. 2i). Therefore, this statistical feature of pairwise connectivity also refines after eye opening, such that a greater proportion of neurons with similar visual responses become bidirectionally connected (Fig. 2h, i).

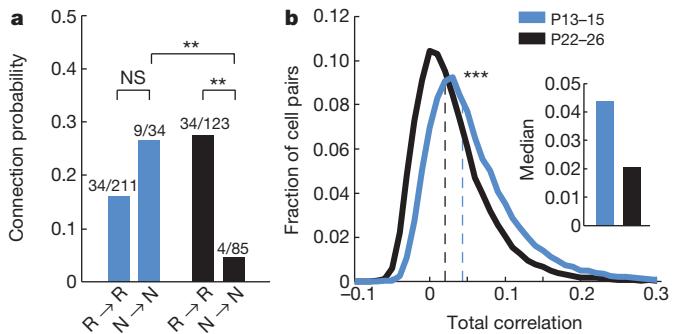
We next tested for developmental changes in recurrent connectivity between neurons not reliably responsive to visual stimuli, which were encountered in similar proportions in both age groups (fraction of neurons not significantly responsive to repeated presentations of the natural movie: P13–15, 39%, 4,133 of 10,509; P22–26, 44%, 4,691 of 10,662). At P22–26, non-responsive neurons connected to each other at much lower rates than responsive neurons ( $P < 0.01$ , Tukey's HSD multiple-comparison test among proportions; Fig. 3a). At P13–15, however, responsive and non-responsive neurons formed recurrent connections at similar rates ( $P > 0.05$ ; Fig. 3a). These data suggest that connections between L2/3 neurons not driven effectively by visual stimuli are selectively eliminated after eye opening ( $P < 0.01$ ; Fig. 3a). This nonspecific-connection scheme at eye opening is consistent with the statistics of population activity during visual stimulation, which showed a twofold-higher total pairwise firing rate correlation at P13–15 than at P22–26 (median correlation: P13–15, 0.044; P22–26, 0.021;  $P < 10^{-307}$ , rank-sum test; Fig. 3b). Higher activity correlations at eye opening may allow connections to be maintained between neurons not reliably driven by visual stimuli. These connections are then lost as activity in the V1 network becomes progressively less correlated.

To obtain a mechanistic insight into the refinement of local recurrent connectivity after eye opening, we constructed a network model of the neocortex based on activity-dependent synaptic plasticity<sup>27</sup>. The model consisted of a recurrently connected cortical network of 18 excitatory and 5 inhibitory integrate-and-fire neurons (Methods). Cortical neurons received feedforward input from 500 presynaptic neurons, a subset of which exhibited spatially correlated activity during each iteration of the simulation (Fig. 4a). The weights of both recurrent and feedforward synapses were updated by a voltage-based spike-timing-dependent plasticity (vSTDP) learning rule<sup>27</sup>. Initially, neurons were seeded with RFs (Methods) and the weights of the recurrent network were drawn randomly from a uniform distribution (Fig. 4b, upper panels). Excitatory neurons in the recurrent network with the same RFs developed strong, mostly bidirectional connections (93.2%; proportions taken across 50 simulations; Fig. 4b, c). There was a strong decrease in connectivity between neurons that were not responsive to feedforward input (from 20.5% near the beginning of the simulation to 0.6% at the end; Fig. 4d) but not between neurons that were both responsive to feedforward input (decrease from 26.0% to 20.7%), consistent with experimental observations (Fig. 3a). To compare the model's behaviour further with our experimental data, we froze the feedforward and recurrent weights at three time points and measured the signal correlation between all responsive neuronal pairs. Higher signal correlations between neurons indicated a higher rate of connectivity at later but not earlier stages of network development (Figs 4e and 2f). The model also predicted the increase in bidirectional connections between neurons with high signal correlations at later stages of development (Figs 4e and 2h). These simulations suggest that feedforward connection patterns determine the structure of recurrent connectivity by activity-dependent mechanisms of synaptic plasticity.



**Figure 2 | Functionally specific connectivity between L2/3 pyramidal cells is not apparent at eye opening.** **a**, Example triplet of neurons shown in a transformed *in vivo* image (left), in the brain slice (middle) and during whole-cell recordings (right). Scale bar, 30  $\mu\text{m}$ . **b**, Membrane potential recordings from neurons shown in **a**. Evoked spikes and average traces of postsynaptic potentials. Dashed lines indicate timing of presynaptic spikes. Vertical scale bars, 80 mV for traces along the diagonal, 0.8 mV otherwise. **c**, Peristimulus time histogram of spikes inferred from calcium signals (arbitrary units) of the three neurons in response to a natural movie sequence (averages of six repetitions). **d**, Schematics of synaptic connectivity and *in vivo* signal correlations during natural movies for the three neurons in **a**. **e**, Overall connectivity rates at eye opening and in more mature V1;  $\chi^2$  test. **f**, Relationship between connection probability and signal correlation of neuronal pairs significantly responsive to the natural movie across age; Cochran-Armitage test. **g**, Relationship between connection probability and difference in preferred orientation ( $\Delta\text{Ori}$ ) among pairs in which both neurons were responsive and orientation selective (orientation selectivity index  $>0.4$ ). **h**, **i**, The probability of observing uni- or bidirectionally connected pairs as a function of either signal correlation (**h**) or  $\Delta\text{Ori}$  (**i**). P13-15,  $n = 13$  mice; P22-26,  $n = 18$  mice.

Our results indicate that RFs exist before mature patterns of recurrent connectivity. However, transient electrical coupling via gap junctions between clonally related neurons contributes to shared feature selectivity and raises the possibility that intracortical connectivity may precede and instruct RF formation<sup>7,9,28</sup>. We extended our network model earlier in time to examine the mechanisms by which gap-junction coupling may influence the emergence of RFs and recurrent connectivity. In this model, early recurrent connectivity in the cortical

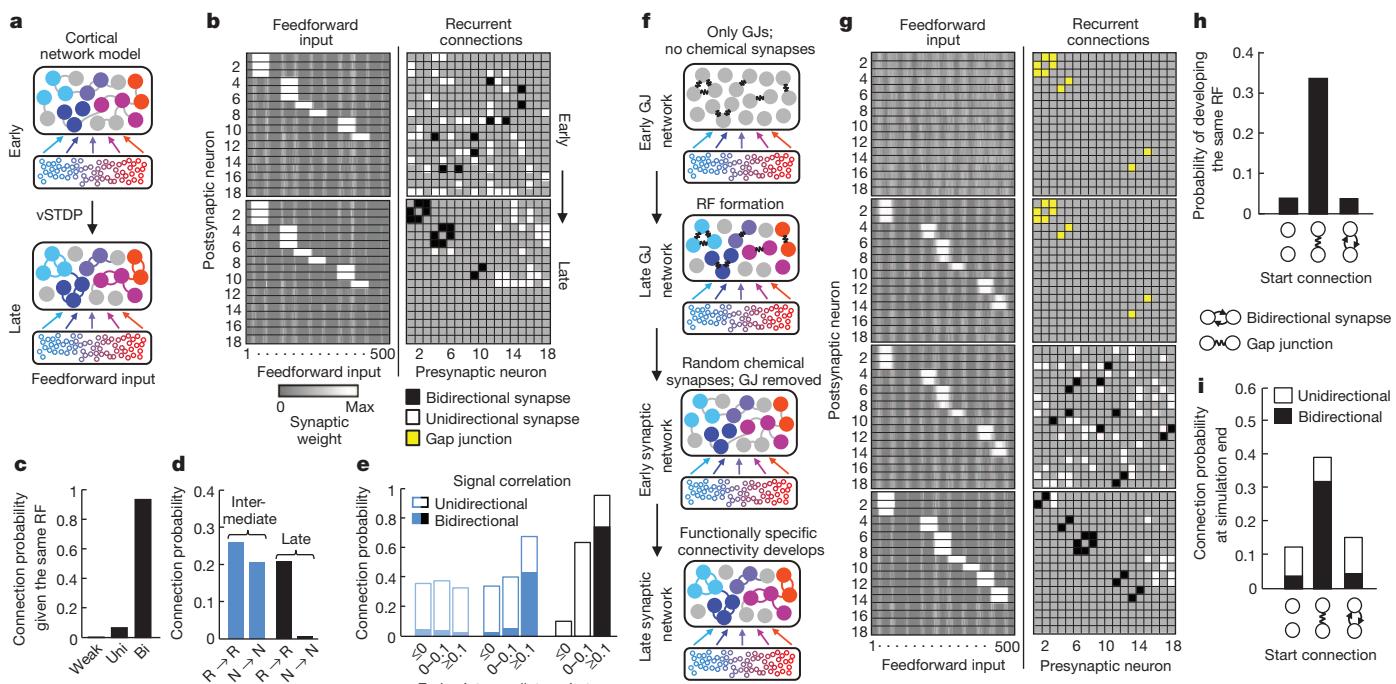


**Figure 3 | Developmental elimination of recurrent connections between non-responsive neurons.** **a**, Connection probability between neurons significantly responsive to the natural movie ( $R \rightarrow R$ ) and between non-responsive neurons ( $N \rightarrow N$ ) at two ages. \*\* $P < 0.01$ , Tukey's HSD multiple-comparison test among proportions; NS, not significant. **b**, Distribution of pairwise time-varying inferred firing rate correlation coefficients for all responsive cell pairs (to natural movies) separated by  $<50 \mu\text{m}$ . \*\*\* $P < 10^{-307}$ , rank-sum test.

network was provided by electrical gap junctions (Fig. 4f, g) and recurrent excitatory chemical synapses did not exist, approximating the organization of the mouse neocortex in the first postnatal week<sup>7</sup>. Early feedforward connections were randomly assigned (Fig. 4f, g) and their weights were updated according to the vSTDP rule. Cell pairs were more likely to stabilize the same set of feedforward inputs (that is, develop the same RF) if they had been connected by gap junctions (gap-junction coupled, 31.6%, 79 of 250 pairs; not gap-junction coupled, 4.1%, 305 of 7,400 pairs;  $P < 0.001$ ,  $\chi^2$  test; 50 simulations; Fig. 4h). We then removed the gap junctions, which disappear by the second postnatal week in mouse V1<sup>7</sup>, and assigned random recurrent synaptic connections to neurons in the cortical network (Fig. 4f, g). Neurons sharing the same RF formed strong synaptic connections (Fig. 4c, g). Therefore, the functional specificity of recurrent connections was influenced indirectly by early gap-junction motifs (Fig. 4i), as electrically coupled neurons were first likely to develop the same RFs before becoming synaptically connected.

Separate simulations initiated with chemical connections revealed that early modifiable bidirectional chemical connections had no influence on either the formation of RFs or recurrent connectivity at the end of the simulation (probability of developing same RF: not connected, 5.5%, 270 of 4,918 pairs; bidirectional chemical connections, 3.8%, 11 of 288 pairs;  $P = 0.22$ ,  $\chi^2$  test; 50 simulations; Fig. 4h, i). Therefore, early initial biases in cortical connectivity may only influence functional circuit development if they exist as strong and stable connections when feedforward inputs are being selected. Although the absolute connectivity rates found experimentally were not perfectly replicated in the model (compare Fig. 2f with Fig. 4e), the connection probability between cells with similar visual responses was higher when assessed deeper in the acute slice, where connections are more likely to be preserved during cutting (supplementary fig. 6 of ref. 5).

We found that the functional specificity of local connections in mouse V1 was not apparent at eye opening despite the occurrence of highly selective responses to visual features. Although the overall rate of synaptic connectivity did not change after eye-opening, connections redistributed according to the following rules: more connections were added preferentially between neurons responding most similarly to visual stimuli, whereas connections were eliminated between cells not reliably responsive to visual stimulation. This result is surprising given existing theories of neural circuit formation, which suggest either that connections are initially exuberant and are subsequently ‘pruned’<sup>10,13</sup> or that synapse number increases after the onset of sensory experience<sup>11,14</sup>. Instead, in local L2/3 networks we find a balanced restructuring of connectivity after eye opening. It remains



**Figure 4 | Feedforward input structure determines the functional organization of recurrent connectivity.** **a**, Sketch of the network model of functional microcircuit development based on vSTDP (see text for details). At the start of the simulation, cortical neurons were randomly connected but received spatially clustered input from a subset of presynaptic neurons. Both feedforward and recurrent connections were updated according to the vSTDP rule (Methods). **b**, Synaptic weight matrices of feedforward (left, reordered for display purposes) and recurrent (right) connections from an example network at the beginning and end of the simulation. Recurrent synaptic connections were classified as weak (light grey), unidirectional (white) or bidirectional (black). **c**, Probability of observing weak, unidirectional or bidirectional connections, at the end of the simulation, between neurons that start with the same RF. **d**, Connection probability of responsive ( $R \rightarrow R$ ) and non-responsive ( $N \rightarrow N$ ) neuronal pairs during and at the end of the simulation. **e**, Relationship

between connection probability and feedforward input-driven signal correlation at three time points in the simulation. **f**, Sketch of different stages of the network model extended to earlier developmental times. GJ, gap junction. **g**, Synaptic weight matrices from the example gap-junction network model. The recurrent network is initially connected with gap junctions (yellow) in the absence of chemical synapses. With time, neurons selected a spatially clustered set of feedforward inputs (RFs). Gap junctions were then removed and recurrent chemical connections were randomly assigned. The simulation was then continued as in **a** and **b**. **h**, Probability of developing RFs from the same set of feedforward inputs for pairs with no recurrent connections, gap junctions or early bidirectional connections (data from separate simulations) at the start of the simulation. **i**, Probability of developing shared connections depended on the starting connectivity. Data in **c–e**, **h** and **i** are from 50 network simulations.

to be seen whether similar mechanisms contribute to the elaboration of long-range connections in visual cortex of rodents and other mammals<sup>22,29,30</sup>.

Functionally specific connection patterns between L2/3 pyramidal cells seem to be instructed by feedforward input (for example from layer 4 or the visual thalamus) only after RF formation. Our network model suggests that correlated firing driven by feedforward activity increases the functional specificity of recurrent connections by activity-dependent mechanisms of synaptic plasticity, which leads to the preferential formation of synapses between any neurons sharing similar RFs. The model can additionally explain how electrically coupled neurons early in development first develop similar feature selectivity and then preferential recurrent connectivity<sup>7,9,28</sup> (Fig. 4g–i).

Our data suggest that functionally organized connectivity between L2/3 pyramidal neurons is not necessary for establishing elemental RF properties at eye opening. Instead, it may contribute to the amplification of visually driven responses and thereby increase the robustness and reliability of cortical representation of sensory input with age (Supplementary Fig. 2c), which may be facilitated by the preferential formation of bidirectional connections between neurons with similar stimulus preferences (Fig. 2h, i). The role of inhibition for the maturation of visual responsiveness remains to be determined.

In conclusion, the patterning of recurrent cortical connectivity through the feedforward-driven, activity-dependent redistribution of connections may be a fundamental rule by which neurons link together into assemblies that process related information.

## METHODS SUMMARY

Responses to different visual stimuli in anaesthetized mouse V1 L2/3 cells were measured using *in vivo* two-photon calcium imaging as previously described<sup>5,25</sup>. Receptive fields were obtained by reverse correlation using a regularized pseudo-inverse method<sup>24</sup>. Synaptic connections were assayed by *in vitro* whole-cell recordings of a subset of neurons imaged *in vivo* and re-identified as previously described<sup>5,25</sup>. The incidence, strength and short-term dynamics of synaptic connections were related to visual response properties across age (Figs 2 and 3 and Supplementary Fig. 4). The network model consisted of 23 recurrently connected integrate-and-fire type neurons (18 excitatory, 5 inhibitory) receiving 500 feedforward inputs with neuronal dynamics and plasticity modelled as in ref. 27.

**Full Methods** and any associated references are available in the online version of the paper.

Received 31 August 2012; accepted 14 February 2013.

1. Binzegger, T., Douglas, R. J. & Martin, K. A. C. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24**, 8441–8453 (2004).
2. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
3. Brown, S. P. & Hestrin, S. Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457**, 1133–1136 (2009).
4. Perin, R., Berger, T. K. & Markram, H. A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. USA* **108**, 5419–5424 (2011).
5. Ko, H. et al. Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
6. Yoshimura, Y., Dantzker, J. L. M. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).
7. Yu, Y.-C. et al. Preferential electrical coupling regulates neocortical lineage-dependent microcircuit assembly. *Nature* **486**, 113–117 (2012).

8. Yuste, R., Peinado, A. & Katz, L. C. Neuronal domains in developing neocortex. *Science* **257**, 665–669 (1992).
9. Li, Y. et al. Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature* **486**, 118–121 (2012).
10. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
11. White, L. E. & Fitzpatrick, D. Vision and cortical map development. *Neuron* **56**, 327–338 (2007).
12. Huberman, A. D., Feller, M. B. & Chapman, B. Mechanisms underlying development of visual maps and receptive fields. *Annu. Rev. Neurosci.* **31**, 479–509 (2008).
13. Innocenti, G. M. & Price, D. J. Exuberance in the development of cortical networks. *Nature Rev. Neurosci.* **6**, 955–965 (2005).
14. Purves, D., White, L. E. & Riddle, D. R. Is neural development Darwinian? *Trends Neurosci.* **19**, 460–464 (1996).
15. Mrsic-Flogel, T. D. et al. Homeostatic regulation of eye-specific responses in visual cortex during ocular dominance plasticity. *Neuron* **54**, 961–972 (2007).
16. Smith, S. L. & Häusser, M. Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nature Neurosci.* **13**, 1144–1149 (2010).
17. Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
18. Hubel, D. H. & Wiesel, T. N. Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol.* **26**, 994–1002 (1963).
19. Chapman, B. & Stryker, M. P. Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.* **13**, 5251–5262 (1993).
20. Krug, K., Akerman, C. J. & Thompson, I. D. Responses of neurons in neonatal cortex and thalamus to patterned visual stimulation through the naturally closed lids. *J. Neurophysiol.* **85**, 1436–1443 (2001).
21. Rochefort, N. L. et al. Development of direction selectivity in mouse cortical neurons. *Neuron* **71**, 425–432 (2011).
22. White, L. E., Coppola, D. M. & Fitzpatrick, D. The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature* **411**, 1049–1052 (2001).
23. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl Acad. Sci. USA* **100**, 7319–7324 (2003).
24. Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D. & Tolhurst, D. J. The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J. Neurosci.* **23**, 4746–4759 (2003).
25. Hofer, S. B. et al. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neurosci.* **14**, 1045–1052 (2011).
26. Reid, R. C. & Alonso, J. M. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **378**, 281–284 (1995).
27. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neurosci.* **13**, 344–352 (2010).
28. Ohtsuki, G. et al. Similarity of visual selectivity among clonally related neurons in visual cortex. *Neuron* **75**, 65–72 (2012).
29. Callaway, E. M. & Katz, L. C. Emergence and refinement of clustered horizontal connections in cat striate cortex. *J. Neurosci.* **10**, 1134–1153 (1990).
30. Ruthazer, E. S. & Stryker, M. P. The role of activity in the development of long-range horizontal connections in area 17 of the ferret. *J. Neurosci.* **16**, 7253–7269 (1996).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. Akerman, D. Attwell, R. Froemke, M. Häusser, C. Levelt, C. Lohmann, T. Margrie, J. Sjöström and members of the Mrsic-Flogel laboratory for advice and comments on the manuscript. We thank D. Farquharson, D. Halpin, A. Hogben of the University College London machine shop for custom parts. The work was supported by the Wellcome Trust (T.D.M.-F., S.B.H., C.B.), the Medical Research Council (L.C.), the European Research Council and the 7th Framework of European Commission ‘EuroV1sion’ grant (T.D.M.-F.), and the Swiss National Science Foundation (C.C., grant no. PA00P3\_139703).

**Author Contributions** H.K. and S.B.H. performed the *in vivo* and *in vitro* experiments and data analysis. S.B.H. performed the RF mapping experiments, and L.C. analysed the RF data with help from J.A. L.C., C.B. and C.C. extended the network model originally developed by C.C. H.K., L.C., S.B.H. and T.D.M.-F. wrote the manuscript. All authors discussed the data and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.B.H. ([s.hofer@ucl.ac.uk](mailto:s.hofer@ucl.ac.uk)) or T.D.M.-F. ([t.mrsic-flogel@ucl.ac.uk](mailto:t.mrsic-flogel@ucl.ac.uk)).

## METHODS

**Animals and surgical procedures.** All experimental procedures were carried out in accordance with institutional animal welfare guidelines and licensed by the UK Home Office. Experiments were performed on C57Bl/6 mice aged P13–15 and P22–35. Mice were initially anaesthetized with a mixture of fentanyl (0.05 mg kg<sup>-1</sup>), midazolam (5.0 mg kg<sup>-1</sup>), and medetomidine (0.5 mg kg<sup>-1</sup>). At the time of imaging, the injectable anaesthetic had mostly worn off and light anaesthesia was maintained by isoflurane (0.3–0.5%) in a 3:2 mixture of O<sub>2</sub>:N<sub>2</sub>O delivered via a small nose cone. Surgery was performed as described previously<sup>15</sup>. Briefly, a small craniotomy (1–2 mm) was carried out over primary visual cortex and sealed after dye injection with 1.6% agarose in HEPES-buffered artificial cerebrospinal fluid (ACSF) and a cover slip.

**In vivo two-photon calcium imaging.** For bulk loading of cortical neurons, the calcium-sensitive dye Oregon Green BAPTA-1 AM (OGB-1 AM, Molecular Probes) was first dissolved in 4 µl DMSO containing 20% Pluronic F-127 (Molecular Probes), and further diluted (1/11) in dye buffer (150 mM NaCl, 2.5 mM KCl and 10 mM HEPES, pH 7.4) to yield a final concentration of 0.9 mM. Sulphorhodamine 101 (SR 101, 50 µM, Molecular Probes) was added to the solution to distinguish neurons and astrocytes<sup>31</sup>. The dye was slowly pressure-injected into the monocular region of right visual cortex at a depth of 170–200 µm with a micropipette (3–5 MΩ, 3–10 p.s.i., 2–4 min) under visual control by two-photon imaging ( $\times 10$  water immersion objective, Olympus). Activity of cortical neurons was monitored by imaging fluorescence changes with a custom-built microscope and a mode-locked Ti:sapphire laser (Mai Tai, Spectra-Physics) at a wavelength of 830 or 930 nm through a  $\times 40$  water immersion objective (0.8 NA, Olympus). Scanning and image acquisition were controlled by custom software written in LABVIEW (National Instruments).

Visual stimuli were generated using the MATLAB (Mathworks) Psychophysics Toolbox<sup>32,33</sup>, and were displayed on an LCD monitor (60-Hz refresh rate) positioned 20 cm from the left eye, roughly at 45° to the long axis of the animal, covering  $\sim 105^\circ \times 85^\circ$  degrees of visual space. At the beginning of each experiment, the appropriate retinotopic position in visual cortex was determined using small grating stimuli at 12–24 neighbouring positions. The monitor was repositioned such that the preferred retinotopic position of most imaged neurons was roughly in the middle of the monitor.

Imaging frames of 256  $\times$  256 pixels were acquired at 7.6 Hz while different visual stimuli, including naturalistic images and movies, and drifting gratings (see below) were presented. After each recording, the focal plane and imaging position was checked and realigned with the initial image if necessary. In combined *in vivo* functional imaging and *in vitro* connectivity assaying experiments, to obtain visually evoked responses from all neurons in a cortical volume of approximately 285 µm  $\times$  285 µm  $\times$  40–120 µm, images were recorded at 7 to 18 cortical depths with a spacing of 7 µm, starting at  $\sim 110$  µm below cortical surface, a depth that corresponds to superficial L2/3 in mouse V1.

Image sequences were aligned for tangential drift and analysed with custom programs written in MATLAB and LABVIEW. Recordings with significant brain movements, vertical drift or both were excluded from further analysis. Cell outlines were detected using a semi-automated algorithm based on morphological measurements of cell intensity, size and shape, and were subsequently confirmed by visual inspection. After erosion of the cell-based regions of interest (to minimize influence of the neuropil signal around the cell bodies), all pixels within each region of interest were averaged to give a single time course ( $\Delta F/F$ ), which was additionally high-pass-filtered at a cut-off frequency of 0.02 Hz to remove slow fluctuations in the signal. Spike trains were inferred from calcium signals using a fast non-negative deconvolution method which approximates the maximum a-posteriori spike train for each neuron, given the fluorescence observations<sup>34</sup>. This method yields spike probabilities (or inferred firing rate) linearly related to the number of action potentials per imaging frame<sup>25</sup>.

**Receptive field measurement.** Receptive field data were acquired from four mice at eye opening (P14–15) and five mature mice (P28–35). Naturalistic image sequences (between 1,440 and 2,700 individual images) were presented on the monitor during two-photon calcium imaging. Images were shown at 2-s intervals (0.5-s presentation time, interleaved by a 1.5-s grey screen) for a total presentation time of between 0.83 and 1.5 h. After the onset of each natural image, 15 imaging frames were recorded at 7.6 Hz before the next image was presented. For each cell in the imaged region, the response to an image was calculated in the following way. Spike probabilities were inferred from calcium signals using the fast non-negative deconvolution method described above. For each visual stimulus,  $k = 1, \dots, N$ , and each cell,  $i = 1, \dots, C$ , the response to the stimulus can be expressed as  $r(k, i, j)$  where  $j = 1, \dots, 15$  are the 15 imaging frames. An average population response was calculated:  $R(j) = \sum_{k,i} r(k, i, j)/NC$ . If the imaging frame,  $J$ , denotes the frame during which the peak average population response occurred (such that  $R(J) = \max_j\{R(j)\}$ ), then the response of cell  $i$  to stimulus  $k$  was defined as  $\sum_{j=J-1}^{J+1} r(k, i, j)/3$ .

To estimate linear RFs, a regularized pseudo-inverse method<sup>24</sup> was used for reverse correlating neuronal responses with images of natural scenes. This algorithm regularizes the inverse problem by introducing a two-dimensional smoothness constraint on the linear RF; namely, the constraint is that the Laplacian of the RF should be close to zero at all points ( $\nabla^2 RF = 0$ ). This method introduces a regularization parameter,  $\lambda$ , which balances the emphasis to be placed on fitting the data and the emphasis to be placed on the smoothness constraint.

Because this method introduces this free parameter, we performed the following analysis to choose the regularization parameter. For each cell and each regularization parameter, the naturalistic images and associated responses were separated into training (75% of the data) and test (remaining 25% of the data) data sets. Training data sets were chosen randomly and the remaining 25% of the data was placed into the test data set. Linear RFs were then calculated using the training data, and a sigmoid nonlinearity, described by the equation

$$P(x) = \frac{A}{1 + \exp(-\alpha x + \beta)}$$

(where  $A$  is the amplitude,  $\alpha$  determines the slope and  $\beta$  determines the offset of the sigmoid), was then fitted to the training data to convert the linear predictions made by the RF into neuronal spike probabilities. Response predictions to the naturalistic images of the test data set were then made and the correlation coefficient between the actual and predicted responses was taken as a measure of RF prediction performance. This procedure was carried out for each cell and each regularization parameter 100 times. For each cell, the regularization parameter that maximized the RF prediction performance was chosen.

To assess whether the RF for a particular cell was significant, the response vector to the naturalistic image sequence was randomly shuffled and the reverse correlation was performed again, using the same regularization parameter,  $\lambda$ . This procedure was repeated 99 times to produce 100 shuffled RFs,  $RF_{shuffled}$ . From these shuffled RFs, the mean,  $\mu_{shuffled}$ , and standard deviation,  $\sigma_{shuffled}$ , across all pixels were calculated. An RF was defined to be significant if there were pixels that had absolute values greater than  $\mu_{shuffled} + 6\sigma_{shuffled}$ .

To fit the Gabor function, we used only the RFs determined to be significant by the previous analysis. The RF was parameterized in MATLAB by fitting a two-dimensional Gabor function using the Levenberg–Marquardt algorithm. The Gabor function is described by

$$G(x, y) = A \exp\left(-\frac{x'^2}{2\sigma_x^2} - \frac{y'^2}{2\sigma_y^2}\right) \cos(2\pi f x' + \varphi)$$

where

$$x' = (x - c_x) \cos(\theta) - (y - c_y) \sin(\theta)$$

$$y' = (x - c_x) \sin(\theta) + (y - c_y) \cos(\theta)$$

These equations describe an underlying two-dimensional cosine grating parameterized by  $\theta$  (orientation),  $f$  (spatial frequency) and  $\varphi$  (phase), which is enveloped by a two-dimensional Gaussian function parameterized by  $A$  (amplitude),  $c_x$  and  $c_y$  (centre of the Gaussian), and  $\sigma_x$  and  $\sigma_y$  (standard deviations of the Gaussian perpendicular to and parallel to the axis of the grating, respectively). Gabor fits were individually inspected to make sure they matched the RF (some Gabor fits were excluded at this point because they did not provide a good match to the RF: P14–15, 3 of 191 (1.6%); P28–35, 6 of 201 (3%)).

To quantify the shapes of RFs, the dimensionless measures  $n_x = \sigma_x/f$  and  $n_y = \sigma_y/f$  were used<sup>35</sup>. These values express the size of the Gaussian envelope in terms of the wavelength of the underlying cosine grating. For instance,  $n_x = 1$  indicates that the standard deviation of the Gaussian perpendicular to the grating is equal to half a cycle of the underlying cosine grating. To get a measure of the size of the RF (Fig. 1 and Supplementary Fig. 1), the visual angle subtended by the Gabor fit along the axis perpendicular to the direction of the cosine grating was calculated. That is, if the eye of the mouse is at (0, 0, 0) in space, the centre of the monitor is at (0, 0,  $d$ ) (where  $d$  is the shortest distance to the mouse eye from the screen), the centre of the Gabor fit to the RF is at ( $c_x, c_y, d$ ), and the angle of orientation of the cosine grating on the screen is  $\theta$ , then the visual angle,  $\alpha$ , subtended by the Gabor was calculated as

$$\alpha = \arccos\left(\frac{u \cdot v}{\|u\| \|v\|}\right)$$

where  $u = (c_x - \sigma_x \cos(\theta), c_y + \sigma_y \sin(\theta), d)$  and  $v = (c_x + \sigma_x \cos(\theta), c_y - \sigma_y \sin(\theta), d)$ . Receptive field similarity was calculated as the pixel–pixel correlation coefficient.

**Natural movie signal correlation.** Natural movies consisted of 40-s sequences of either moving scenes in a mouse cage or compilations of The Life of Mammals (BBC), adjusted to 70% mean contrast, continuously looped six times. Visual

responsiveness to natural movies was determined by the following procedure. For all stimulus repetitions, inferred spike trains were moving-average-filtered with a time window of three frames ( $\sim 0.394$  s). The smoothed firing rates at corresponding points of the stimulus were then treated as groups and tested for differences by one-way analysis of variance. Neurons with a  $P$  value less than 0.01 (that is, those that exhibited consistently increased firing during at least one period of stimulus presentation) were considered significantly visually responsive. For pairs of significantly responsive neurons, the signal correlation was calculated as the Pearson's correlation coefficient of the averaged responses to the stimulus.

**Orientation tuning.** To measure the orientation preference and selectivity of neurons, square-wave gratings (0.035 cycles per degree, 2 cycles per second, 100% contrast) drifting in eight different directions were randomly interleaved, with the grating standing still for 1.4–1.9 s before moving for 0.9–1.5 s (six repetitions per grating). Responsive neurons that exhibited consistently increased firing during at least one time point of presentation of each grating were identified by one-way analysis of variance. Among cells responsive to grating stimuli ( $P < 0.05$ ), the mean inferred firing rate during the presentation of a drifting grating was taken as the response to each stimulus. From each trial, we obtained one orientation tuning curve, and neurons were defined as reliably responsive if the mean cross-correlation between all pairs of curves obtained from different trials was greater than 0.1. Responses from different trials were then averaged to obtain the average orientation tuning curve for each neuron. This orientation tuning curve was then Fourier interpolated to 360 points, and the preferred direction was determined by the angle at which the interpolated tuning curve attained its maximum. The preferred orientation was taken as preferred direction modulo 180°. Orientation selectivity index (OSI) was calculated as  $(R_{\text{best}} - R_{\text{ortho}})/(R_{\text{best}} + R_{\text{ortho}})$ , where  $R_{\text{best}}$  is the interpolated response to the preferred direction and  $R_{\text{ortho}}$  is the average of interpolated responses to the directions orthogonal to the best-responding direction. When relating connection probability to orientation selectivity, neurons were defined as orientation selective if OSI  $> 0.4$ . To quantify neuronal response reliability, we calculated the coefficient of variation (s.d. divided by mean) from responses to the optimal grating direction.

**In vitro whole-cell recording.** We carried out imaging experiments followed by whole-cell recordings *in vitro* at P13–15 and P22–26, using an approach described previously<sup>5</sup>. After two-photon calcium imaging of visual responses *in vivo*, small volumes of red fluorescent microspheres (Lumafluor) were injected into the imaged region to facilitate identification of the region in the slice tissue. The mouse brain was then rapidly removed to and dissected in ice-cold ACSF containing 125 mM NaCl, 2.5 mM KCl, 1 mM MgCl<sub>2</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 2 mM CaCl<sub>2</sub>, 26 mM NaHCO<sub>3</sub>, 25 mM dextrose; 315–325 mOsm, bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub>, pH 7.4. Visual cortex slices (300  $\mu$ m) were cut coronally (HM 650 V Vibration Microtome, MICROM) and were incubated at 34 °C for 30 min before they were transferred to the recording chamber. The slice containing the imaged region was identified by the presence of OGB-1 green fluorescence and the red microsphere injection site. To reveal the relative locations of cells, a detailed morphological stack of the slice was acquired with a custom-built microscope and a mode-locked Ti:sapphire laser (Chameleon, Coherent), at a wavelength of 830 nm, through a  $\times 16$  water immersion objective (0.8 NA, Nikon). Scanning and image acquisition were controlled by custom software written in LABVIEW.

Whole-cell recordings from two to six cells were carried out in regions identified by visually comparing image stacks obtained *in vivo* and *in vitro*, using red fluorescent microspheres and the pial surface as reference. Recordings were carried out in ACSF at 28 °C, using Multiclamp 700B amplifiers (Axon Instruments), and data was acquired using custom software running in IGOR PRO<sup>36</sup> (WaveMetrics Inc.) or MATLAB. Recording pipettes were filled with internal solution containing 5 mM KCl, 115 mM K-gluconate, 10 mM K-HEPES, 4 mM MgATP, 0.3 mM NaGTP, 10 mM Na-phosphocreatine, 0.1% w/v biocytin, 40  $\mu$ M Alexa Fluor 594; 290–295 mOsm, pH 7.2. Junction potential was not corrected for. The chloride reversal potential was about –85.2 mV. Cells were approached under visual guidance using laser-scanning Dodt contrast. After breakthrough, the presence of synaptic connections was tested by evoking five spikes at 30 Hz in each cell, repeated 30 to 120 times, while searching for postsynaptic responses.

The paired-pulse ratio (PPR) was calculated as the amplitude of the second evoked excitatory postsynaptic potential (EPSP) divided by that of the first one. After connectivity mapping, step currents from 250 pA to 700 pA were injected in 50-pA increments if the input resistance was smaller than around 400 MΩ, and, if necessary (as indicated by early ceasing of firing due to inactivation of voltage-gated sodium channels), currents from –125 pA to 350 pA were injected in 25-pA increments for neurons with larger input resistance. Pyramidal neurons were identified from several parameters: their morphology in Alexa-594-filled image stacks (Fig. 2a); regular-spiking pattern on current injection; spike half-width ( $> 1$  ms); and, in the presence of connections, depolarizing postsynaptic potentials (Fig. 2b). To match the same neurons imaged *in vivo* and recorded from *in vitro*,

we performed three-dimensional image registration of *in vivo* and *in vitro* image stacks by affine transformation using custom-written MATLAB software subsequent to the experiment.

Connection probabilities were calculated as the number of connections detected over the number of potential connections assayed. Probabilities of unidirectional and bidirectional connections were calculated as the numbers of unidirectionally and, respectively, bidirectionally connected pairs divided by the total number of pairs. To relate connectivity to functional properties, the asymptotic Cochran-Armitage test for trend was used to test for the significance of linear trends<sup>37</sup>. Pairs in which a high-quality recording was achieved in only one cell (for example when the other cell was too depolarized or unhealthy, or the seal resistance was less than 1 GΩ) connectivity was assayed only in the direction from the unhealthy cell to the healthy cell, given that spikes could be evoked in both cells. Data from these pairs were included in the analysis of connection probability, but not in the analysis of the probability of finding bidirectional or unidirectional pairs. Only neuronal pairs in which both neurons were located at  $> 60$   $\mu$ m depth from the slice surface and with an intersoma distance of  $< 50$   $\mu$ m were included in the analysis.

The strength and short-term plasticity of synapses were also measured because these synaptic parameters affect the efficacy of presynaptic firing on postsynaptic partners. In P13–15 mice, connections between L2/3 pyramidal cells were significantly stronger (median EPSP amplitude: P13–15, 0.41 mV; P22–26, 0.20 mV;  $P = 2.9 \times 10^{-4}$ , rank-sum test; Supplementary Fig. 4a) and PPR was significantly lower (median PPR: P13–15, 0.87; P22–26, 1.13;  $P = 6.2 \times 10^{-4}$ , rank-sum test; Supplementary Fig. 4b) than in P22–26 mice, in line with previous reports<sup>38</sup>. However, in neither age group was there significant correlation between EPSP amplitude or PPR, on the one hand, and signal correlation or difference in preferred orientation, on the other (Supplementary Fig. 4c–f). Part of the connectivity data was published previously in ref. 5.

**Neuron model.** In the network model, the dynamics of the membrane potential,  $u(t)$ , of model neurons is described by the adaptive exponential integrate-and-fire model<sup>39</sup>:

$$C \frac{du}{dt} = -g_L(u - E_L) + g_L A_T e^{(u - V_T)/A_T} - w_{\text{ad}} + z + I$$

Here  $C$  is the membrane capacitance,  $g_L$  is the leak conductance,  $E_L$  is the resting potential and  $I$  is the stimulating current. The exponential term describes the activation of sodium current. The parameter  $A_T$  is called the slope factor and  $V_T$  is the threshold potential. A hyperpolarizing adaptation current is described by the variable  $w_{\text{ad}}$ , which has dynamics

$$\tau_{w_{\text{ad}}} \frac{dw_{\text{ad}}}{dt} = a(u - E_L) - w_{\text{ad}}$$

where  $\tau_{w_{\text{ad}}}$  is the time constant of the adaption of the neuron and  $a$  is a parameter. Upon firing, the variable  $u$  is reset to a fixed value,  $V_{\text{reset}}$ , whereas  $w_{\text{ad}}$  is increased by an amount  $b$ . An additional current,  $z$ , which is set to a value  $I_{\text{sp}}$  immediately after a spike occurs and otherwise decays with a time constant  $\tau_z$  such that

$$\tau_z \frac{dz}{dt} = -z$$

was used to account for spike afterpotential<sup>40</sup>. Refractoriness is modelled with the adaptive threshold,  $V_T$ , which starts at  $V_{T_{\text{max}}}$  after a spike and decays to  $V_{T_{\text{rest}}}$  with a time constant  $\tau_{V_T}$ , such that

$$\tau_{V_T} \frac{dV_T}{dt} = -(V_T - V_{T_{\text{rest}}})$$

Parameters for the neuron model were taken from ref. 39, and  $\tau_z$  was set to 40 ms in agreement with refs 36, 41 and was kept fixed throughout all simulations. We made the following choices:  $C = 281$  pF,  $g_L = 30$  nS,  $E_L = -70.6$  mV,  $A_T = 2$  mV,  $V_{T_{\text{rest}}} = -50.4$  mV,  $\tau_{w_{\text{ad}}} = 144$  ms,  $a = 4$  nS,  $b = 0.0805$  pA,  $I_{\text{sp}} = 400$  pA,  $\tau_z = 40$  ms,  $\tau_{V_T} = 50$  ms,  $V_{T_{\text{max}}} = -30.4$  mV.

**Plasticity model.** Our plasticity model exhibits separate additive contributions to the plasticity rule, one for long-term depression (LTD) and another one for long-term potentiation<sup>42</sup> (LTP). Synaptic weights had hard bounds imposed at  $w_{\text{min}}$  and  $w_{\text{max}}$ . For the LTD part, we assumed that presynaptic spike arrival at synapse  $i$  induces the depression of the synaptic weight  $w_i$  by an amount  $-A_{\text{LTD}}[\bar{u}_-(t) - \theta_-]_+$ . The brackets  $[...]$  indicate rectification, that is, any value  $\bar{u}_- < \theta_-$  does not lead to a change<sup>43</sup>. The quantity  $\bar{u}_-(t)$  is an exponential low-pass-filtered version of the postsynaptic membrane potential,  $u(t)$ , with time constant  $\tau_-$ , such that

$$\tau_- \frac{d\bar{u}_-(t)}{dt} = -\bar{u}_-(t) + u(t)$$

Because the presynaptic spike train is described as a series of short pulses at time  $t_i^n$ ,  $X_i(t) = \sum_n \delta(t - t_i^n)$ , where  $i$  is the index of the synapse and  $n$  is the index that

counts the spike, the depression—that is, the change in  $w_i$  due to LTD (indicated by the superscript minus sign)—is given by

$$\frac{dw_i^-}{dt} = -A_{\text{LTD}}(\bar{u})X_i(t)[\bar{u}_- - \theta_-]_+$$

if  $w_i > w_{\min}$ , where  $A_{\text{LTD}}(\bar{u})$  is an amplitude parameter that is under the control of homeostatic processes<sup>44</sup> and  $\bar{u}$  is the running average of depolarization of the postsynaptic neuron, averaged over a time scale of 1 s. The time scale of 1 s is not critical (100 s or more would be more realistic for homeostasis), but is convenient for the numerical implementation.

For the LTP part, we assumed that each presynaptic spike at the synapse  $w_i$  increases the trace,  $\bar{x}_i(t)$ , of some biophysical quantity, which decays exponentially with a time constant  $\tau_x$  (refs 45, 46), such that

$$\tau_x \frac{d\bar{x}_i(t)}{dt} = -\bar{x}_i(t) + X_i(t)$$

where  $X_i(t)$  is the spike train defined above. Potentiation—the change in  $w_i$  due to LTP (indicated by the superscript plus sign)—is given by

$$\frac{dw_i^+}{dt} = A_{\text{LTP}}\bar{x}_i(t)[u - \theta_+]_+[\bar{u}_+ - \theta_-]_+$$

if  $w_i < w_{\max}$ . Here  $A_{\text{LTP}}$  is a free amplitude parameter fitted to electrophysiology data<sup>27</sup> and  $\bar{u}_+(t)$  is another low-pass-filtered version of  $u(t)$ , similar to  $\bar{u}_-(t)$  but with a shorter time constant,  $\tau_+$ . Thus, positive weight changes can occur if the momentary voltage,  $u(t)$ , surpasses a threshold,  $\theta_+$ , and, at the same time, the average value,  $\bar{u}_+(t)$ , is greater than  $\theta_-$ .

The final rule used in the simulation was

$$\begin{aligned} \frac{dw_i}{dt} = & -A_{\text{LTD}}(\bar{u})X_i(t)[\bar{u}_- - \theta_-]_+ \\ & + A_{\text{LTP}}\bar{x}_i(t)[u - \theta_+]_+[\bar{u}_+ - \theta_-]_+ \end{aligned}$$

combined with hard bounds  $w_{\min} \leq w_i \leq w_{\max}$ . For network simulations,  $A_{\text{LTD}}(\bar{u}) = A_{\text{LTD}}\bar{u}^2/u_{\text{ref}}^2$ , where  $u_{\text{ref}}^2$  is a reference value. We made the following parameter choices for simulations:  $\theta_- = -70.6$  mV,  $\theta_+ = -45.3$  mV,  $A_{\text{LTD}} = 14 \times 10^{-5}$  mV<sup>-1</sup>,  $A_{\text{LTP}} = 8 \times 10^{-5}$  mV<sup>-2</sup>,  $\tau_x = 15$  ms,  $\tau_- = 10$  ms,  $\tau_+ = 7$  ms.

**Network simulation.** In all simulations, 500 presynaptic Poisson neurons with firing rates  $v_i^{\text{pre}}$  ( $i = 1, \dots, 500$ ) were connected to 18 postsynaptic excitatory neurons. The input rates  $v_i^{\text{pre}}$  followed a Gaussian profile, that is,  $v_i^{\text{pre}} = A \exp(-(i-\mu)^2/2\sigma^2)$ , with variance  $\sigma = 10$  and amplitude  $A = 30$ . The centre,  $\mu$ , of the Gaussian shifted randomly every 100 ms between ten equally distributed positions, each position occurring with equal probability. Circular boundary conditions were assumed, that is, neuron  $i = 500$  was considered to neighbour neuron  $i = 1$ . Five inhibitory neurons were each driven by 14 excitatory neurons and projected onto 11 excitatory neurons. These connections were chosen randomly and were fixed with a weight equal to 1. Feedforward connections onto the inhibitory neurons were drawn from a uniform distribution on the interval [0, 0.5] and were fixed for the duration of the simulation. The reference value was set to  $u_{\text{ref}}^2 = 70$  mV<sup>2</sup>. Parameters for the feedforward connections were chosen as for the plasticity model. The excitatory recurrent connections were plastic under the same rule and with the same parameters as the feedforward connections, but the amplitudes  $A_{\text{LTP}}$  and  $A_{\text{LTD}}$  were reduced by a factor of 100.

In the first set of simulations, feedforward weights were initialized with RF (weights taken from previous test simulations took values between 0 and 3, which were also the hard bounds). At the beginning of the simulation, recurrent excitatory connection weights were drawn randomly from a uniform distribution on the interval [0, 0.75] (hard bounds were set to 0 and 0.75). At each time point during the simulation, noise current (Gaussian white noise) was independently injected into each cell in the recurrent network. In this and all subsequent networks, the model was run for 20 s of simulated time to allow the homeostatic dynamical variables to settle before the recurrent synaptic weights were reinitialized. The simulations were then run for another 1,000 s.

To calculate signal correlations in the network at three different time points (at the reinitialization of recurrent excitatory chemical synapses, 1 s of simulation

time later and at the end of the simulation), the weights were frozen and the same stimuli were played to assess the firing rate correlations across neurons. Two neurons were considered to be bidirectionally connected if both synaptic weights between them were  $>0.6$ , and were considered to be unidirectionally connected if only one of the synaptic weights was  $>0.6$ .

Responsiveness was determined in the following way: the feedforward weights onto each neuron were summed, to produce a single value for each cell. These values were then plotted on a histogram, which displayed a bimodal distribution. A threshold value was chosen between the two peaks of this bimodal distribution, to separate the cells into responsive and non-responsive. Feedforward weights were manually checked to make sure no cells with clear RFs were missed. Neuronal pairs whose RFs had a correlation coefficient of  $>0.85$  were defined to have the same RF.

The recurrent gap-junction network began such that neurons 1 and 2, 3 and 4, and 5 to 7 were electrically coupled together in the stated combinations, and was run in this state for 200 s. There were no chemical synapses during this time. The gap-junction model was taken from ref. 47. The current from neuron  $i$  to neuron  $j$  was defined as  $I_{ij}(t) = \alpha \sum_i \delta(t - t_i) - g_{\text{gap}}[u_j(t) - u_i(t)]$ , where we chose  $g_{\text{gap}} = 2$  nS and chose  $\alpha$  such that the spikelet was about 2 mV. The network was in this state for 200 s of simulation time (after the initial 20-s settling time), at which point gap junctions were removed and replaced with recurrent excitatory chemical connections with weights drawn randomly from a uniform distribution on the interval [0, 0.75]. After this time point, Gaussian white noise current stimulation was again provided to each cell in the recurrent network. The network was then run for another 800 s.

The comparison chemical network was run in the same manner as the gap-junction network except that a recurrent chemical network, with weights drawn randomly from a uniform distribution on the interval [0, 0.75], replaced the recurrent gap-junction network during the first 200 s, and there was no replacement of the recurrent weights after this 200 s.

31. Nimmerjahn, A., Kirchhoff, F., Kerr, J. N. D. & Helmchen, F. Sulforhodamine 101 as a specific marker of astroglia in the neocortex *in vivo*. *Nature Methods* **1**, 31–37 (2004).
32. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
33. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
34. Vogelstein, J. T. et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–3704 (2010).
35. Ringach, D. L. Mapping receptive fields in primary visual cortex. *J. Physiol. (Lond.)* **558**, 717–728 (2004).
36. Sjöström, P. J., Turrigiano, G. G. & Nelson, S. B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
37. Agresti, A. *Categorical Data Analysis* 2nd edn 181–182 (Wiley-Interscience, 2002).
38. Frick, A., Feldmeyer, D. & Sakmann, B. Postnatal development of synaptic transmission in local networks of L5A pyramidal neurons in rat somatosensory cortex. *J. Physiol. (Lond.)* **585**, 103–116 (2007).
39. Brette, R. & Gerstner, W. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.* **94**, 3637–3642 (2005).
40. Shouval, H. Z., Bear, M. F. & Cooper, L. N. A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *Proc. Natl. Acad. Sci. USA* **99**, 10831–10836 (2002).
41. Badel, L. et al. Dynamic I-V curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J. Neurophysiol.* **99**, 656–666 (2008).
42. O'Connor, D. H., Wittenberg, G. M. & Wang, S.-H. Dissection of bidirectional synaptic plasticity into saturable unidirectional processes. *J. Neurophysiol.* **94**, 1565–1573 (2005).
43. Artola, A., Bröcher, S. & Singer, W. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature* **347**, 69–72 (1990).
44. Turrigiano, G. G. & Nelson, S. B. Homeostatic plasticity in the developing nervous system. *Nature Rev. Neurosci.* **5**, 97–107 (2004).
45. Gerstner, W., Kempter, R., Van Hemmen, J. L. & Wagner, H. A neuronal learning rule for sub-millisecond temporal coding. *Nature* **383**, 76–78 (1996).
46. Pfister, J.-P. & Gerstner, W. Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neurosci.* **26**, 9673–9682 (2006).
47. Ostojic, S., Brunel, N. & Hakim, V. Synchronization properties of networks of electrically coupled neurons in the presence of noise and heterogeneities. *J. Comput. Neurosci.* **26**, 369–392 (2009).

# Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway

Jaekyoung Son<sup>1\*</sup>, Costas A. Lyssiotis<sup>2,3†\*</sup>, Haoqiang Ying<sup>4</sup>, Xiaoxu Wang<sup>1</sup>, Sujun Hua<sup>4</sup>, Matteo Ligorio<sup>5</sup>, Rushika M. Perera<sup>6</sup>, Cristina R. Ferrone<sup>5</sup>, Edouard Mullarky<sup>2,3†</sup>, Ng Shyh-Chang<sup>2,7</sup>, Ya'an Kang<sup>8</sup>, Jason B. Fleming<sup>8</sup>, Nabeel Bardeesy<sup>6</sup>, John M. Asara<sup>3,9</sup>, Marcia C. Haigis<sup>10</sup>, Ronald A. DePinho<sup>4</sup>, Lewis C. Cantley<sup>2,3†</sup> & Alec C. Kimmelman<sup>1</sup>

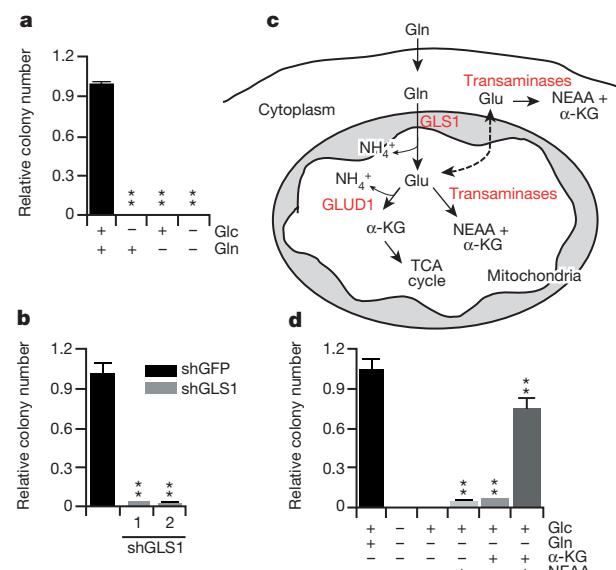
Cancer cells have metabolic dependencies that distinguish them from their normal counterparts<sup>1</sup>. Among these dependencies is an increased use of the amino acid glutamine to fuel anabolic processes<sup>2</sup>. Indeed, the spectrum of glutamine-dependent tumours and the mechanisms whereby glutamine supports cancer metabolism remain areas of active investigation. Here we report the identification of a non-canonical pathway of glutamine use in human pancreatic ductal adenocarcinoma (PDAC) cells that is required for tumour growth. Whereas most cells use glutamate dehydrogenase (GLUD1) to convert glutamine-derived glutamate into  $\alpha$ -ketoglutarate in the mitochondria to fuel the tricarboxylic acid cycle, PDAC relies on a distinct pathway in which glutamine-derived aspartate is transported into the cytoplasm where it can be converted into oxaloacetate by aspartate transaminase (GOT1). Subsequently, this oxaloacetate is converted into malate and then pyruvate, ostensibly increasing the NADPH/NADP<sup>+</sup> ratio which can potentially maintain the cellular redox state. Importantly, PDAC cells are strongly dependent on this series of reactions, as glutamine deprivation or genetic inhibition of any enzyme in this pathway leads to an increase in reactive oxygen species and a reduction in reduced glutathione. Moreover, knockdown of any component enzyme in this series of reactions also results in a pronounced suppression of PDAC growth *in vitro* and *in vivo*. Furthermore, we establish that the reprogramming of glutamine metabolism is mediated by oncogenic KRAS, the signature genetic alteration in PDAC, through the transcriptional upregulation and repression of key metabolic enzymes in this pathway. The essentiality of this pathway in PDAC and the fact that it is dispensable in normal cells may provide novel therapeutic approaches to treat these refractory tumours.

The prognosis of patients with PDAC remains dismal. The disease is extremely aggressive and is profoundly resistant to all forms of therapy<sup>3</sup>. Thus, there is a strong impetus to identify new therapeutic targets for this cancer. In recent years, there has been renewed interest in understanding the altered metabolism in cancer, and how such dependencies can be targeted for therapeutic gain. However, achieving a successful therapeutic index remains a major challenge to the development of effective cancer therapies that target metabolic pathways.

Recent evidence demonstrates that some cancer cells use glutamine (Gln) to support anabolic processes that fuel proliferation<sup>2</sup>. However, the importance of Gln metabolism in pancreatic tumour maintenance is not known. Thus, we sought to explore the dependence of PDAC on Gln, and to examine the functional role of Gln in PDAC metabolism. As expected from our previous work<sup>4</sup>, glucose was required for PDAC

growth. Additionally, PDAC cells were also profoundly sensitive to Gln deprivation, indicating that Gln is also critical for PDAC growth (Fig. 1a and Supplementary Fig. 1).

Gln provides a carbon source to fuel the tricarboxylic acid (TCA) cycle and nitrogen for nucleotide, nonessential amino acid (NEAA) and hexosamine biosynthesis<sup>5,6</sup>. To assess the role of Gln metabolism in PDAC growth, we first impaired glutaminase (GLS) activity using RNA interference (RNAi). Notably, GLS knockdown markedly reduced PDAC growth (Fig. 1b and Supplementary Fig. 2a, b). Consistent with this observation, glutamate (Glu) was able to support growth in Gln-free conditions (Supplementary Fig. 2c).



**Figure 1 | PDAC use a non-canonical glutamine (Gln) metabolism pathway.** **a**, PDAC proliferation requires both glucose and Gln. Cells were plated in complete media (10 mM glucose and 2 mM Gln) which was replaced the following day with glucose- or Gln-free medium supplemented with 10% dialysed fetal bovine serum. **b**, Relative clonogenic growth of 8988T cells expressing a control shRNA (shGFP) or two independent shRNAs to GLS1. **c**, Schematic overview of GLUD1- or transaminase-mediated Glu metabolism. **d**, Relative clonogenic growth of 8988T cells.  $\alpha$ -KG (4 mM), NEAA mixture (0.1 mM glycine, alanine, aspartate, asparagine, proline and serine) or the combination was added to media following Gln withdrawal.  $\alpha$ -KG,  $\alpha$ -ketoglutarate; Glc, glucose; Gln, glutamine; Glu, glutamate; NEAA, non-essential amino acid. Error bars represent s.d. of triplicate wells from a representative experiment. \*\* $P$  < 0.01.

<sup>1</sup>Division of Genomic Stability and DNA Repair, Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>2</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>3</sup>Division of Signal Transduction, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. <sup>4</sup>Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>5</sup>Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>6</sup>Cancer Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>7</sup>Stem Cell Transplantation Program, Stem Cell Program, Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Dana Farber Cancer Institute, Boston, Massachusetts 02130, USA. <sup>8</sup>Department of Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>9</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. <sup>10</sup>Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>†</sup>Present address: Department of Medicine, Weill Cornell Medical College, New York, New York 10065, USA.

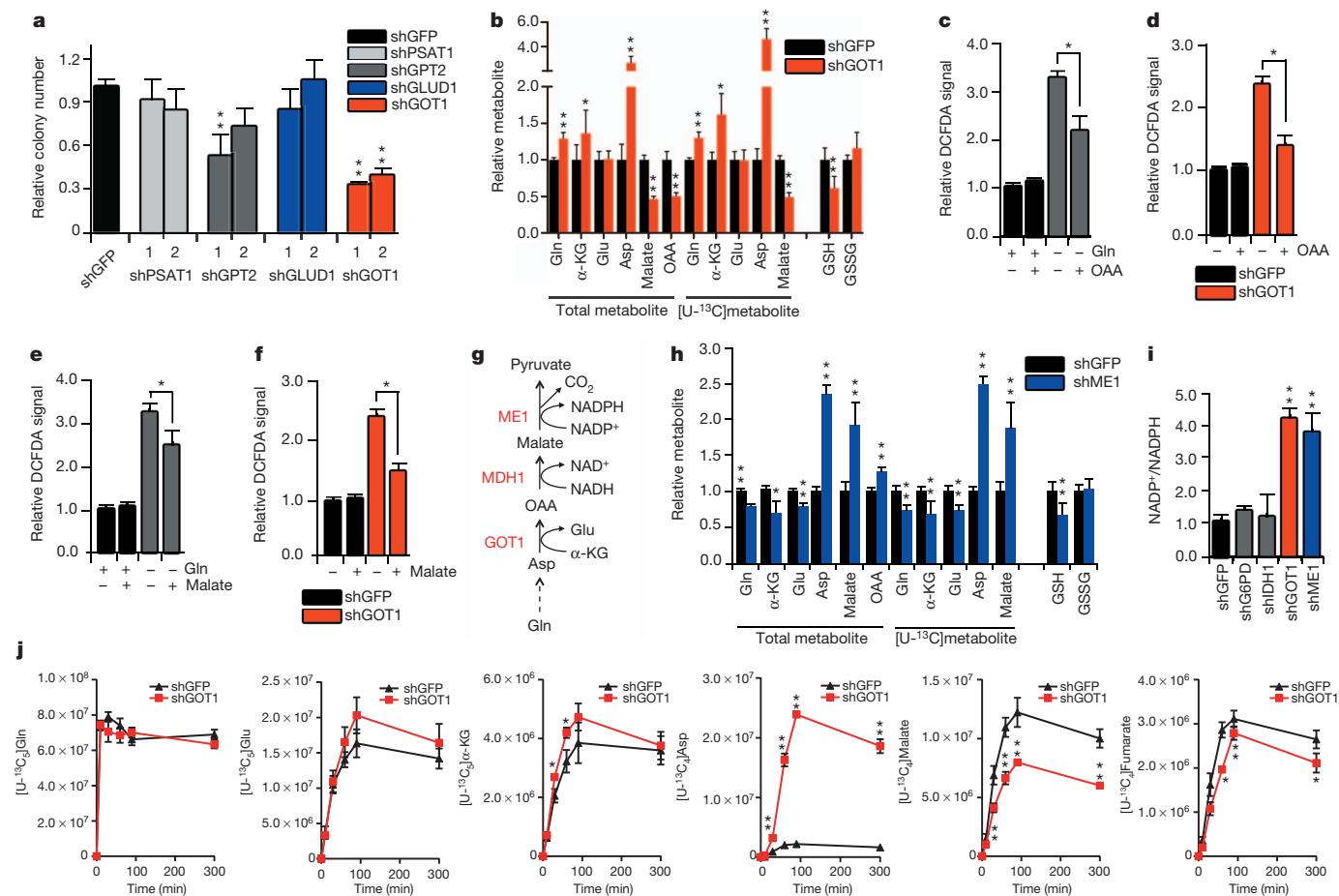
\*These authors contributed equally to this work.

Glu can be converted into  $\alpha$ -ketoglutarate ( $\alpha$ -KG) to replenish the TCA cycle through two mechanisms<sup>1</sup>; either by glutamate dehydrogenase (GLUD1) or transaminases (Fig. 1c). Indeed, many cancer cells rely on GLUD1-mediated Glu deamination to fuel the TCA cycle<sup>7</sup>, and  $\alpha$ -KG has been shown to be an essential metabolite in Gln metabolism<sup>8</sup>. Surprisingly, dimethyl  $\alpha$ -KG<sup>9</sup> did not restore growth upon Gln deprivation (Fig. 1d), whereas the combination of  $\alpha$ -KG and an NEAA mixture (the output of transaminase-mediated Glu metabolism) markedly rescued proliferation in multiple PDAC lines (Fig. 1d and Supplementary Fig. 2d, e). Together, these data indicate that PDAC cells metabolize Gln in a manner that is different from canonical models<sup>10</sup> and that transaminases may be critical for Gln metabolism in PDAC.

To confirm the importance of transaminases in PDAC Gln metabolism, we treated PDAC cells with either aminoxyacetate (AOA), a pan-inhibitor of transaminases<sup>11</sup>, or epigallocatechin gallate (EGCG), an inhibitor of GLUD1<sup>12</sup>. Whereas EGCG had no effect on PDAC growth, AOA treatment robustly inhibited the growth of multiple PDAC cell lines (Supplementary Fig. 3). Consistent with these results, GLUD1 knockdown also had no effect on PDAC growth (Fig. 2a). To identify

the specific transaminase(s) involved in PDAC Gln metabolism, we inhibited a panel of Glu-dependent transaminases (aspartate, alanine and phosphoserine transaminase) individually using RNAi and examined the effect on PDAC growth. Among this panel, knockdown of the aspartate transaminase GOT1 significantly impaired PDAC growth in multiple PDAC cell lines and primary PDAC cells (Fig. 2a and Supplementary Fig. 4, 5).

We next explored the direct effects of GOT1 on Gln metabolism by performing targeted metabolomic analysis in GOT1 knockdown PDAC cells using uniformly  $^{13}\text{C}$ -labelled Gln ( $[\text{U}-^{13}\text{C}_5]\text{Gln}$ ) as a tracer<sup>4,13</sup>. GOT1 catalyses the conversion of aspartate (Asp) and  $\alpha$ -KG into oxaloacetate (OAA) and Glu in the cytoplasm. Indeed, GOT1 knockdown led to increased Gln-derived Asp (and total Asp) and decreased OAA (Fig. 2b and Supplementary Fig. 6a). Interestingly, we also observed a significant decrease in the ratio of reduced-to-oxidized glutathione (GSH:GSSG; Fig. 2b and Supplementary Fig. 6b), suggesting that GOT1 may play a role in the maintenance of cellular redox homeostasis. It should be noted that the changes in metabolite abundance described in this experiment are representative of the total cellular metabolite pool



**Figure 2 | GOT1 is essential for redox balance and growth in PDAC.**

**a**, Relative clonogenic growth of 8988T cells expressing a control shRNA (shGFP) or two independent shRNAs targeting GLUD1, GOT1, GPT2 or PSAT1. Error bars represent s.d. of triplicate wells from a representative experiment. **b**, Relative metabolite abundance in 8988T cells grown in  $[\text{U}-^{13}\text{C}_5]\text{Gln}$  upon GOT1 knockdown. Data are presented as the total metabolite pool (encompassing both metabolite derived from Gln and that not Gln-derived, left) and the  $^{13}\text{C}$ -labelled, Gln-derived metabolite pool (right). Error bars represent the s.d. of three independently prepared samples. **c-f**, Relative ROS levels in 8988T cells under conditions indicated as determined by DCFDA (2',7'-dichlorodihydrofluorescein diacetate) staining. DCFDA assay was performed 24 h after supplementing Gln-free media with either OAA (4 mM) or dimethyl

malate (4 mM). Each bar represents the mean of three independent experiments with error bars representing the s.d. **g**, Schematic depiction of the cytoplasmic reactions that convert Asp into pyruvate. **h**, Relative metabolite abundance in 8988T cells grown in  $[\text{U}-^{13}\text{C}_5]\text{Gln}$  upon ME1 knockdown. **i**, NADP<sup>+</sup>/NADPH ratio in 8988T cells expressing a control shRNA (shGFP), or an shRNA to G6PD, IDH1, GOT1 or ME1. Error bars represent s.d. of six replicate wells from a representative experiment. **j**, Flux of the Gln carbon skeleton into downstream metabolites as a function of time. The reads for uniformly  $^{13}\text{C}$ -labelled metabolites, presented in ion current, are plotted for cells expressing the shGFP control or shRNA to GOT1. Asp, aspartate; OAA, oxaloacetate; Glu, glutamate; GSH, reduced glutathione; GSSG, oxidized glutathione. Error bars represent the s.d. of three independently prepared samples. \* $P < 0.05$ ; \*\* $P < 0.01$ .

due to technical limitations associated with organelle-specific metabolite isolation. Importantly however, the results we obtained are consistent with what one would expect if flux through GOT1 was impaired.

In PDAC, we recently demonstrated that oncogenic KRAS enhances the flux of glycolytic intermediates specifically through the non-oxidative arm of the pentose phosphate pathway (PPP) to support DNA/RNA biosynthesis without affecting the NADPH-producing oxidative arm<sup>4</sup>. This decoupling of ribose biogenesis from NADPH production by KRAS suggests that PDAC cells may rely on an alternative mechanism to maintain cellular redox balance. Indeed, inhibition of the oxidative PPP in PDAC by knockdown of the rate-limiting enzyme glucose-6-phosphate dehydrogenase (G6PD) had minimal effect on reactive oxygen species (ROS) levels and, consistent with this, glucose deprivation had only a modest impact on ROS (Supplementary Fig. 6c). Therefore, we speculated that the GOT1-mediated conversion of Gln-derived Asp into OAA functions in a pathway that is used to generate the NADPH which could be used to maintain redox balance. To test this hypothesis, we assessed ROS levels upon Gln deprivation in the absence or presence of OAA. Indeed, Gln deprivation induced ROS, and OAA could partially rescue the elevated ROS levels (Fig. 2c). GOT1 knockdown also increased ROS levels, which again were significantly restored upon supplementation with OAA (Fig. 2d).

Given that Gln-derived malate (and total malate) was significantly reduced upon GOT1 knockdown (Fig. 2b), we suspected that Gln-derived OAA is metabolized into malate, which is used by malic enzyme (ME1) to create NADPH, providing the reducing power to maintain reduced glutathione pools<sup>14</sup>. Indeed, malate was able to partially rescue the oxidative stress imposed by Gln-deprivation (Fig. 2e) or GOT1 knockdown (Fig. 2f). These data are consistent with a model whereby Gln-derived Asp is converted by GOT1 into OAA, then converted into malate by malate dehydrogenase (MDH1) and subsequently oxidized by ME1 into pyruvate and reducing power in the form of NADPH (Fig. 2g). Consistent with this pathway, metabolomic analysis of [ $U-^{13}\text{C}_5$ ]Gln tracing in ME1 knockdown cells revealed a significant increase in Asp, malate and OAA and decreased GSH (Fig. 2h and Supplementary Fig. 6d). Furthermore, knockdown of GOT1 and ME1 markedly increased the cellular NADP<sup>+</sup>/NADPH ratio (Fig. 2i), whereas inhibition of other cytosolic sources of NADPH (G6PD or isocitrate dehydrogenase, IDH1) had no effect on NADP<sup>+</sup>/NADPH ratios or ROS (Fig. 2i and Supplementary Fig. 6c). Together, the data suggest that PDAC use Gln through the pathway depicted in Fig. 2g to increase the NADPH/NADP<sup>+</sup> ratio for maintenance of redox homeostasis. Lastly, Gln tracing kinetic flux experiments in GOT1 knockdown cells clearly demonstrate decreased flux through this pathway (Fig. 2j and Supplementary Fig. 7). Interestingly, lactate in the  $^{13}\text{C}$ -tracing experiments was minimally labelled, indicating that the pyruvate produced by ME1 is not used to make lactate by lactate dehydrogenase and has a yet undetermined fate.

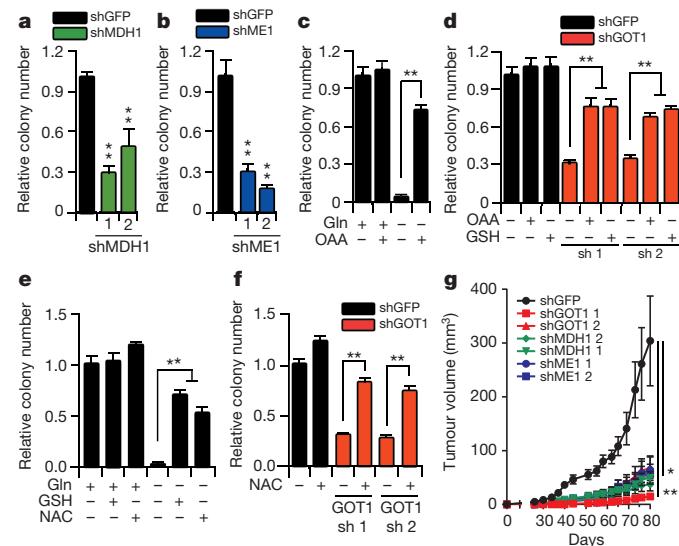
The majority of Asp in PDAC cells (50–75%) is derived from Gln, as evidenced by  $^{13}\text{C}$ -labelling (Supplementary Fig. 6a, d and 8a). In principle, uniformly  $^{13}\text{C}$ -labelled Asp can be derived from Gln following either (1) the GLUD1-mediated conversion of Glu to  $\alpha$ -KG (and its subsequent traversing through the TCA cycle) or (2) the mitochondrial aspartate transaminase (GOT2)-mediated conversion of Glu and OAA to  $\alpha$ -KG and Asp. Of these two enzymes, only GOT2 knockdown significantly affected PDAC growth (Fig. 2a and Supplementary Fig. 8b). Consistent with this observation, GLUD1 knockdown did not affect Asp biosynthesis from Gln, whereas GOT2 knockdown resulted in a significant decrease in Gln-derived Asp in PDAC cells (Supplementary Fig. 8a, c).

We next tested whether other components of this pathway are also necessary to support PDAC growth. Indeed, knockdown of either MDH1 or ME1 markedly inhibited clonogenic survival of PDAC cells (Fig. 3a, b and Supplementary Fig. 9) in a manner similar to GOT1 knockdown. As a further test of enzyme knockdown specificity, we investigated the ability of central metabolites in this pathway to rescue

PDAC growth upon Gln deprivation. First, we confirmed that exogenously added metabolites permeated the cell and populated metabolite pools (Supplementary Fig. 10). Next, we demonstrated that the combination of GOT1 substrates, Asp and  $\alpha$ -KG, could rescue cell growth in Gln-free conditions (Supplementary Fig. 11a). Additionally, OAA permitted PDAC growth under Gln-free conditions in multiple PDAC cell lines (Fig. 3c and Supplementary Fig. 11) as well as upon both GLS (Supplementary Fig. 12) and GOT1 knockdown (Fig. 3d and Supplementary Fig. 13a, b). Lastly, the addition of dimethyl-malate<sup>15</sup> was able to partially rescue PDAC cell growth upon Gln deprivation (Supplementary Fig. 13c) or GOT1 knockdown (Supplementary Fig. 13d).

Next, to confirm that the OAA or malate rescue of PDAC growth upon Gln deprivation was through maintenance of redox homeostasis, we treated cells grown in Gln-free conditions with a cell-permeable GSH analogue. Remarkably, GSH dramatically rescued clonogenic growth following Gln-deprivation (Fig. 3e and Supplementary Fig. 11c) or GOT1 knockdown (Fig. 3d and Supplementary Fig. 13a, b). GSH was also able to rescue either MDH1 or ME1 knockdown (Supplementary Fig. 14a). Near-identical results were also seen with the antioxidant N-acetylcysteine (NAC) (Fig. 3e, f and Supplementary Fig. 14b, c). Together, these data support the idea that Gln is used by PDAC cells to maintain redox homeostasis, which is required to support cellular proliferation.

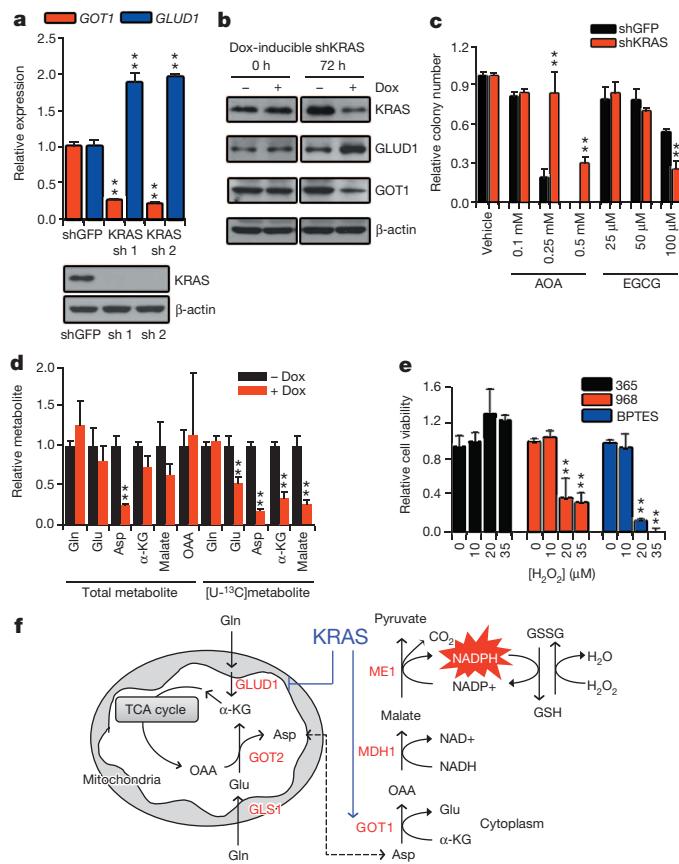
As further confirmation of the importance of this pathway in PDAC, we suppressed GOT1, MDH1 and ME1 expression using two lentiviral short hairpin RNAs (shRNAs) in PDAC cells and assessed their ability to grow as xenografts. Consistent with our *in vitro* results, both GLUD1 shRNAs had no effect on tumour growth (Supplementary Fig. 15). In contrast, GOT1, MDH1 and ME1 knockdown each robustly diminished tumour growth (Fig. 3g). These data provide further support for the critical role of this pathway in Gln metabolism and PDAC tumour growth.



**Figure 3 | Metabolism of the Gln carbon skeleton through GOT1, MDH1 and ME1 supports PDAC growth by maintaining redox balance.**

**a, b,** Relative clonogenic growth of 8988T cells expressing a control shRNA (shGFP) or two independent shRNAs to MDH1 or ME1. **c, e,** Relative clonogenic growth of 8988T cells under conditions indicated. Cells were plated in complete culture media (10 mM glucose and 2 mM Gln), which was replaced the following day with Gln-free medium supplemented with OAA (4 mM), GSH (4 mM) or N-acetylcysteine (NAC) (4 mM). **d, f,** Relative clonogenic growth of 8988T cells expressing a control shRNA (shGFP) or two independent shRNAs to GOT1 with or without OAA (4 mM), GSH (4 mM) or NAC (4 mM). Error bars represent the s.d. of triplicate wells from a representative experiment (**a–f**). **g,** Xenograft growth of 8988T cells expressing a control shRNA (shGFP), shRNAs to GOT1 (1 and 2), MDH1 (1 and 2) or ME1 (1 and 2) in mice. Error bars represent s.e.m. ( $n = 10$ ). \* $P < 0.05$ ; \*\* $P < 0.01$ .

In contrast to PDAC, this pathway seems to be dispensable in normal cells. Indeed, treatment of non-transformed human pancreatic ductal cells (HPDEs) and human diploid fibroblasts (IMR90) with AOA had only modest effects on growth (Supplementary Fig. 16a). HPDE cells, unlike PDAC cells, were significantly sensitive to EGCG, suggesting a greater reliance on the activity of GLUD1 (Supplementary Fig. 16b). Consistent with these results, GOT1 knockdown did not impair the growth of HPDE and IMR90 (Supplementary Fig. 16c). We also obtained similar results in mouse ductal epithelial cells (mPDE) and mouse PDAC cell lines, with the mPDEs being highly insensitive to AOA and GOT1 knockdown and two independently derived mouse PDAC lines showing significant sensitivity to AOA and GOT1 knockdown (Supplementary Fig. 16d, e). Furthermore, using an inducible shGOT1 construct, we demonstrated that mouse PDAC rely on GOT1 to sustain tumour growth *in vivo* (Supplementary



**Figure 4 | Oncogenic KRAS mediates Gln reprogramming in PDAC.**

**a**, Expression of GLUD1 and GOT1 was determined by quantitative RT-PCR in 8988T cells expressing a control shRNA (shGFP) or two independent shRNAs targeting KRAS. Western blot confirmed knockdown of KRAS expression. Error bars represent the s.d. of three replicates. **b**, The effect of KRAS knockdown on GLUD1 or GOT1 protein levels in Panc1 cells expressing a doxycycline-inducible KRAS shRNA. **c**, Relative clonogenic growth of 8988T cells, expressing a control shRNA (shGFP) or shRNA to KRAS, following treatment with AOA or EGCG. Error bars represent the s.d. of triplicate wells from a representative experiment. **d**, Relative metabolite abundance in Miapaca2 cells grown in [<sup>13</sup>C]-Gln following doxycycline-inducible KRAS knockdown. Data are presented as the total metabolite pool (encompassing both metabolite derived from Gln and that not Gln-derived, left) and the uniformly <sup>13</sup>C-labelled, Gln-derived metabolite pool (right). Error bars represent the s.d. of three independently prepared samples. **e**, Relative cell viability of 8988T cells treated with GLS inhibitor 968 (active) (10 μM), 365 (inactive analogue) (50 μM) or BPTES (bis-2-(5-phenylacetamido-1,2,4-thiadiazol-2-yl)ethyl sulfide) (100 nM) with increasing concentrations of H<sub>2</sub>O<sub>2</sub>. Error bars represent the s.d. of triplicate wells from a representative experiment. **f**, Model depicting the KRAS-regulated Gln metabolic reprogramming in PDAC used to maintain redox and support growth. AOA, aminoxyacetate; EGCG, epigallocatechin gallate. \*P < 0.05; \*\*P < 0.01.

Fig. 17). Collectively, these data demonstrate that the GOT1-mediated usage of the Gln carbon skeleton is a metabolic adaptation that PDAC, and not normal cells, have uniquely acquired to support growth.

Our previous work demonstrated that anabolic glucose metabolism in PDAC is controlled by oncogenic KRAS, which leads to altered expression of a number of rate-limiting metabolic enzymes<sup>4</sup>. To investigate the role of KRAS in the reprogramming of Gln metabolism, we assessed the expression of GOT1 and GLUD1 upon knockdown of KRAS in PDAC cells. Interestingly, KRAS knockdown resulted in a marked increase in GLUD1 and a decrease in GOT1 expression at the transcriptional level (Fig. 4a), as well as the protein level (Fig. 4b) in multiple PDAC lines (Supplementary Fig. 18a). Additionally, using five independent orthotopic tumours derived from our inducible KRAS PDAC model<sup>4</sup>, we show that expression of GOT1 increased and GLUD1 decreased in an oncogenic KRAS-dependent manner *in vivo* (Supplementary Fig. 18b). These findings demonstrate that, in PDAC, oncogenic KRAS plays a critical role in coordinating the shift in Gln metabolism to maintain tumour growth and survival.

We next assessed the sensitivity of PDAC cells to either AOA or EGCG upon KRAS knockdown using 8988T cells, a cell line that is not dependent on KRAS for survival<sup>16</sup>. Consistent with our previous results, AOA significantly inhibited clonogenic growth, whereas EGCG had minimal effects. Interestingly, KRAS knockdown made the cells significantly more resistant to AOA and sensitive to EGCG (Fig. 4c and Supplementary Fig. 18c). To confirm the role of oncogenic KRAS in the reprogramming of Gln metabolism, a targeted metabolomic analysis using [<sup>13</sup>C]-Gln was performed upon KRAS knockdown. Indeed, the changes observed were consistent with KRAS supporting the anabolic metabolism of Gln, where multiple metabolites in the GOT1-dependent pathway were significantly deregulated (Fig. 4d and Supplementary Fig. 19).

Given the importance of Gln metabolism in maintaining the redox state of PDAC, we speculated that inhibition of anabolic Gln metabolism may sensitize PDAC to oxidative stress. To test this concept, we inhibited Gln metabolism in PDAC cells using a GLS inhibitor and examined whether this would synergize with hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) treatment. First, we found that two chemically distinct GLS inhibitors<sup>8,17</sup> had a growth-suppressive effect on both human and mouse PDAC cells (Supplementary Fig. 20a, b), consistent with the GLS knockdown data (Fig. 1b). In addition, when combined with H<sub>2</sub>O<sub>2</sub>, this effect was notably augmented, indicating that PDAC cells are markedly more sensitive to ROS when Gln metabolism is impaired (Fig. 4e and Supplementary Fig. 20c). This finding may have significant therapeutic implications, given that clinical grade GLS inhibitors are being developed<sup>18</sup> and that standard PDAC therapies (such as radiation) lead to the generation of ROS. Moreover, because this aspect of Gln metabolism does not seem as critical in normal cells, these data indicate an accessible therapeutic window.

Collectively, our data reveal a novel dependence on the transaminases GOT2 and GOT1 for metabolism of the glutamine carbon skeleton in PDAC. These reactions lead to the cytosolic conversion of aspartate into OAA, malate and then pyruvate and are required to sustain PDAC growth, probably through maintaining redox balance. Importantly, our work also demonstrates that oncogenic KRAS mediates this reprogramming of glutamine metabolism (Fig. 4f). Although this pathway is critical for redox balance and cell growth in PDAC, it does not preclude the involvement of other pathways that may contribute to redox balance such as glutathione synthesis<sup>9,19,20</sup> or an NRF2-dependent mechanism<sup>21</sup>. Furthermore, there may be cell-type-specific differences, as KRAS-transformed fibroblasts require both GLUD1 and transaminases for cell growth<sup>22</sup>. Lastly, these findings may have implications for future therapeutic approaches as inhibition of glutamine metabolism in PDAC can potentially synergize with therapies that increase intracellular reactive oxygen species, such as chemotherapy and radiation.

## METHODS SUMMARY

Proliferation and clonogenic assays were performed as previously described<sup>23</sup>. To characterize Gln metabolism, targeted liquid chromatography-tandem mass spectrometry was performed<sup>4</sup>. Briefly, cells were grown in complete media and transferred into Gln-free media supplemented with [ $U\text{-}^{13}\text{C}_5$ ]Gln overnight (steady state) or for the indicated time points (flux analyses). For subcutaneous xenografts, PDAC cells infected with lentiviral shRNAs to suppress target gene expression were suspended in 100  $\mu\text{l}$  HBSS and injected subcutaneously into the lower flank of NCr nude mice. For mouse xenografts, murine PDAC cells stably infected with a doxycycline-inducible GOT1 shRNA construct were injected. Animals were fed with doxycycline water starting on the day of injection or when tumour volume reached  $\sim 50 \text{ mm}^3$ . Complete images of western blots are presented in Supplementary Fig. 21.

**Full Methods** and any associated references are available in the online version of the paper.

Received 15 May 2012; accepted 21 February 2013.

Published online 27 March 2013.

1. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B.. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
2. Wise, D. R. & Thompson, C. B. Glutamine addiction: a new therapeutic target in cancer. *Trends Biochem. Sci.* **35**, 427–433 (2010).
3. Hidalgo, M. Pancreatic cancer. *N. Engl. J. Med.* **362**, 1605–1617 (2010).
4. Ying, H. et al. Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell* **149**, 656–670 (2012).
5. Dang, C. V. Links between metabolism and cancer. *Genes Dev.* **26**, 877–890 (2012).
6. DeBerardinis, R. J. & Cheng, T. Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene* **29**, 313–324 (2010).
7. Shanware, N. P., Mullen, A. R., DeBerardinis, R. J. & Abraham, R. T. Glutamine: pleiotropic roles in tumor growth and stress resistance. *J. Mol. Med.* **89**, 229–236 (2011).
8. Wang, J. B. et al. Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell* **18**, 207–219 (2010).
9. Weinberg, F. et al. Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. *Proc. Natl Acad. Sci. USA* **107**, 8788–8793 (2010).
10. Jones, R. G. & Thompson, C. B. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev.* **23**, 537–548 (2009).
11. Wise, D. R. et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA* **105**, 18782–18787 (2008).
12. Choo, A. Y. et al. Glucose addiction of TSC null cells is caused by failed mTORC1-dependent balancing of metabolic demand with supply. *Mol. Cell* **38**, 487–499 (2010).
13. Yuan, M., Breitkopf, S. B., Yang, X. & Asara, J. M. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nature Protocols* **7**, 872–881 (2012).

14. Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nature Rev. Cancer* **11**, 85–95 (2011).
15. Huypens, P. et al. The dicarboxylate carrier plays a role in mitochondrial malate transport and in the regulation of glucose-stimulated insulin secretion from rat pancreatic beta cells. *Diabetologia* **54**, 135–145 (2011).
16. Singh, A. et al. A gene expression signature associated with “K-Ras addiction” reveals regulators of EMT and tumor cell survival. *Cancer Cell* **15**, 489–500 (2009).
17. DelaBarre, B. et al. Full-length human glutaminase in complex with an allosteric inhibitor. *Biochemistry* **50**, 10764–10770 (2011).
18. Vander Heiden, M. G. Targeting cancer metabolism: a therapeutic window opens. *Nature Rev. Drug Discov.* **10**, 671–684 (2011).
19. Trachootham, D. et al. Selective killing of oncogenically transformed cells through a ROS-mediated mechanism by beta-phenylethyl isothiocyanate. *Cancer Cell* **10**, 241–252 (2006).
20. DeBerardinis, R. J. et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl Acad. Sci. USA* **104**, 19345–19350 (2007).
21. DeNicola, G. M. et al. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106–109 (2011).
22. Gaglio, D. et al. Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. *Mol. Syst. Biol.* **7**, 523 (2011).
23. Yang, S. et al. Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25**, 717–729 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to thank D. Anastasiou for advice and helpful comments on the manuscript; M. Yuan and S. Breitkopf for technical help with mass spectrometry experiments; J. Erickson, R. Cerione and J. Escobedo for the GLS inhibitors. Grant support derives from the Dana Farber Cancer Institute (A.C.K.), National Cancer Institute Grant R01 CA157490 (A.C.K.), Kimmel Scholar Award (A.C.K.), AACR-PanCAN Career Development Award (A.C.K.), NIH grants T32 CA009382–26 (H.Y.) and P01 CA117969 (L.C.C. and R.A.D.). C.A.L. is the Amgen Fellow of the Damon Runyon Cancer Research Foundation (DRG-2056-10). NIH grants 5P01CA120964-05 and Dana-Farber/Harvard Cancer Center Support Grant 5P30CA006516-46 (J.M.A.).

**Author Contributions** J.S., C.A.L., L.C.C. and A.C.K. designed the study, interpreted the data and wrote the manuscript. J.S., C.A.L., H.Y. and X.W. performed the experiments. J.M.A., E.M. and N.S. helped with the metabolomic studies and with S.H., M.C.H. and R.A.D. assisted in data interpretation. M.L., R.M.P., C.R.F., Y.K., N.B. and J.B.F. developed essential reagents and resources.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests; details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.C.C. ([lec2014@med.cornell.edu](mailto:lec2014@med.cornell.edu)) or A.C.K. ([Alec\\_Kimmelman@DFCI.harvard.edu](mailto:Alec_Kimmelman@DFCI.harvard.edu)).

## METHODS

**Cell culture.** Cell lines were obtained from the American Type Culture Collection or the German Collection of Microorganisms and Cell Cultures. All cell lines were tested routinely, and before all metabolomic analyses, for mycoplasma contamination. RPMI-1640, fetal bovine serum and dialysed fetal bovine serum (dFBS) were purchased from Invitrogen. Glucose-free DMEM (containing 2 mM Gln), dimethyl  $\alpha$ -KG, Asp, GSH reduced ethyl ester, OAA, dimethyl malate and DMEM powder (without glucose and Gln) were obtained from Sigma, and Gln-free RPMI 1640 was purchased from Cellgro. Cosmic calf serum (CCS) was obtained from Thermo Scientific. Cells were cultured in the following media: 8988T, Panc1, MPanc96, Miapaca2 and PL45 in DMEM supplemented with 10 mM glucose and 10% CCS; 8902 in RPMI with 10% CCS; IMR90 in MEM with 10% FBS; HPDE cells were cultured as described previously<sup>24</sup>. Primary human PDAC lines were generated from ascites fluid under IRB approved protocols 02-240 and 2007P001918. Lines were sequenced and confirmed to have KRAS mutations.

**Cell proliferation assay**<sup>23</sup>. Cells were plated in 24-well plates at 2,000 cells per well in 0.5 ml of media. To deprive Gln, cells were plated in complete culture media (10 mM glucose and 2 mM Gln), which was exchanged with Gln-free medium supplemented with 10% dFBS the following day. Media was not changed throughout the course of the experiment. At the indicated time points, cells were fixed in 10% formalin and stained with 0.1% crystal violet. Dye was extracted with 10% acetic acid and the relative proliferation was determined by attenuation (*D*) at 595 nm.

**Clonogenic assay**<sup>23</sup>. Cells were plated in 6-well plates at 300 cells per well in 2 ml of media. Media was not changed throughout the course of the experiment. After 7–10 days, colonies were fixed in 80% methanol and stained with 0.2% crystal violet.

**Quantitative RT-PCR.** Total RNA was extracted using TRIzol (Invitrogen) and reverse transcription was performed from 2  $\mu$ g of total RNA using oligo-dT and MMLV HP reverse transcriptase (Epicentre), according to the manufacturer's instructions. Quantitative RT-PCR was performed with SYBR Green dye using an Mx3000PTM instrument (Stratagene). PCR reactions were performed in triplicate and the relative amount of cDNA was calculated by the comparative  $C_t$  method using the 18S ribosomal RNA sequences as a control. See Supplementary Information for primer sequences.

**Xenograft studies.** For subcutaneous xenografts, 8988T cells were infected with lentiviral shRNAs targeting GLUD1 ( $n = 2$ ), GOT1 ( $n = 2$ ), MDH1 ( $n = 2$ ), ME1 ( $n = 2$ ) and GFP (control hairpin,  $n = 1$ ) and subjected to a short puromycin selection (2  $\mu$ g ml<sup>-1</sup>); shRNA sequences are presented below. 1.5  $\times$  10<sup>6</sup> cells, suspended in 100  $\mu$ l Hanks Buffered Saline Solution (HBSS), were injected subcutaneously into the lower flank of NCr nude mice (Taconic). Tumour length and width were measured twice weekly and the volume was calculated according to the formula (length  $\times$  width<sup>2</sup>)/2. All xenograft experiments with human PDAC lines were approved by the HMS Institutional Animal Care and Use Committee (IACUC) under protocol number 04-605. For mouse xenografts, a doxycycline-inducible GOT1 shRNA construct was generated by annealing oligonucleotides that form an shRNA sequence to mouse GOT1, shRNA (forward: CCGGCCA CATGAGAACGTTCTCGAGAAGAACGTTCTCATGGTT-TTG; reverse: AATTCAAAAACCACATGAGAACGTTCTCGAGA AGAAACGTCTCTCATGTGG). The annealed product was ligated into the AgeI-EcoRI sites of the pLKO-Tet-on vector. For subcutaneous xenograft, 10<sup>6</sup> stably infected murine PDAC cells were suspended in 100  $\mu$ l Hanks Buffered Saline Solution and injected subcutaneously into the lower flank of NCr nude mice (Taconic). Animals were fed with doxycycline water (doxycycline 2 g l<sup>-1</sup>, sucrose 20 g l<sup>-1</sup>) starting on the day of injection or when tumour diameter reached 50 mm. Tumour volumes were measured every third day starting from day 4 post-injection and calculated as above. These xenograft experiments were approved under MDACC IACUC protocol 111113931.

**Western blot analysis.** After SDS-PAGE, proteins were transferred to Hybond-N Nitrocellulose (Amersham Biosciences). Membranes were blocked in Tris-buffered saline (TBS) containing 5% non-fat dry milk and 0.1% Tween 20 (TBS-T), before incubation with the primary antibody overnight at 4 °C. The membranes were then washed with TBS-T followed by exposure to the appropriate horseradish peroxidase-conjugated secondary antibody for 1 h and visualized on Kodak X-ray film using the enhanced chemiluminescence (ECL) detection system (Thermo Scientific). The following antibodies were used: KRAS (F234, Santa Cruz), GOT1 (NBP1-54778, Novus), GLUD1 (ab55061, Abcam) and  $\beta$ -actin (A2066, Sigma).

**ROS quantification.** DCFDA assay was performed 24 h after supplementing Gln-free media with either OAA (4 mM) or dimethyl malate (4 mM). Cells were incubated with 5  $\mu$ M 2',7'-dichlorodihydrofluorescein diacetate (DCFDA, Invitrogen) for 30 min. Excess DCFDA was removed by washing the cells twice with PBS, and labelled cells were then trypsinized, rinsed, and resuspended in PBS. Oxidation of DCFDA to the highly fluorescent 2',7'-dichloro-fluorescein (DCF) is proportionate to ROS generation and was analysed by flow cytometry.

**Metabolomics.** For steady state metabolomic analysis, PDAC cell lines were grown to ~50% confluence in growth media (DMEM, 2 mM Gln, 10 mM glucose, 10% CCS) on 10-cm dishes in biological quadruplicate. A complete media change was performed 2 h before metabolite collection. To trace Gln metabolism, PDAC cell lines were grown as above and then transferred into Gln-free DMEM (with 10 mM glucose) containing 10% dialysed FBS and 2 mM [ $U$ -<sup>13</sup>C<sub>5</sub>]Gln (Cambridge Isotope Labs) overnight (for steady state labelling) or for the indicated time points in the flux analyses. Additionally, fresh media containing [ $U$ -<sup>13</sup>C<sub>5</sub>]Gln was exchanged 2 h before metabolite extraction for steady state analyses. The quantity of the metabolite fraction analysed was adjusted to the corresponding protein concentration calculated upon processing a parallel 10-cm dish. Metabolite fractions were collected and analysed by targeted LC-MS/MS via selected reaction monitoring (SRM), as described<sup>5,14</sup>. Processed data was analysed in Cluster 3.0 and TreeViewer.

**Measurement of sensitization of PDAC cells to ROS.** PDAC cell lines were plated into 96-well plates at 10<sup>3</sup> cells per well in 200  $\mu$ l of growth media. The following day, growth media was replaced with that containing GLS inhibitors and/or H<sub>2</sub>O<sub>2</sub>. Parallel plates were analysed at 3, 6 and 9 days by Cell Titer Glo analysis (Promega), per the manufacturer's instruction. The GLS inhibitors 968 (active) and 365 (structurally similar, inactive) were provided as a gift from the Cerione laboratory<sup>9</sup>. BPTEs was a gift from Jaime Escobedo (Forma Therapeutics).

**Lentiviral-mediated shRNA targets.** All shRNA vectors were obtained from the RNA Interference Screening Facility of Dana Farber Cancer Institute. The sequences and RNAi Consortium clone IDs for each shRNA are as follows. shGLS-1, GCACAGACATGGTTGGTATAT (TRCN0000051135); shGLS-2, GCCCTGAAGCAGTCGAAATA (TRCN0000051136); shMDH1-1, CCCTGT TGTAATCAAGAATAA (TRCN0000221892); shMDH1-2, GCAACAGATAAA AGAAGACGTT (TRCN0000221893); shME1-1, GCCTTCAATGAACGGCC TATT (TRCN0000064728); shME1-2, CCAAACAATATAAGTTGGTGT (TRCN0000064729); shGLUD1-1, CCCAAGAACTATACTGATAAT (TRCN 0000220878); shGLUD1-2, GCAGAGTTCCAAGACAGGATA (TRCN00000220880); shGOT1-1, CGCTTGGTACAATGGAAACAAA (TRCN0000034784); shGOT1-2, GCTAATGACAATAGCCTAAAT (TRCN0000034785); shGPT2-1, CGGCATTCTACGATCCTGAA (TRCN0000035024); shGPT2-2, CCATCAA ATGGCTCCAGACAT (TRCN0000035025); shPSAT1-1, GCCAAGAGTT GGGACTATA (TRCN0000035264); shPSAT1-2, CCAGACAACATAAGGT GATT (TRCN0000035265); shKRAS-1, CCTCGTTCTACACAGAGAAA (TRCN0000040148); shKRAS-2, GAGGGCTTCTTGTGTATT (TRCN000033260).

**Reagents.** NADP<sup>+</sup>/NADPH ratios were determined using the NADP/NADPH assay kit (Abcam; ab65349) according to the manufacturer's instructions. Briefly, 10<sup>5</sup> cells ( $n = 6$  wells of a 6-well dish) were collected on ice in extraction buffer and subject to two rounds of freeze-thaw at -80 °C. NADP<sup>+</sup> and NADPH values were determined in biological sextuplet and concentration was obtained by comparison to standard curves. OAA was not analysed by targeted LC/MS-MS due to its limited stability in aqueous solvents at room temperature<sup>25</sup>. As such, the abundance of this metabolite was determined using a quantification kit (Biovision), according to the manufacturer's instruction. Briefly, 2  $\times$  10<sup>6</sup> cells ( $n = 4$ , 10-cm dishes) were collected during log-phase growth by trypsinization, resuspended immediately in the buffers provided (on ice), analysed and compared to standard curves. The signals obtained were normalized to the protein concentration calculated upon processing a parallel 10-cm dish. NEAA mixture consisted of 0.1 mM glycine, alanine, aspartate, asparagine, proline and serine.

24. Ouyang, H. et al. Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *Am. J. Pathol.* **157**, 1623–1631 (2000).
25. Bajad, S. U. et al. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1125**, 76–88 (2006).

# Proteolytic elimination of N-myristoyl modifications by the *Shigella* virulence factor IpaJ

Nikolay Burnaevskiy<sup>1</sup>, Thomas G. Fox<sup>2</sup>, Daniel A. Plymire<sup>3</sup>, James M. Ertelt<sup>4</sup>, Bethany A. Weigle<sup>1</sup>, Andrey S. Selyunin<sup>1</sup>, Sing Sing Way<sup>4</sup>, Steven M. Patrie<sup>3</sup> & Neal M. Alto<sup>1</sup>

**Protein N-myristoylation is a 14-carbon fatty-acid modification that is conserved across eukaryotic species and occurs on nearly 1% of the cellular proteome<sup>1,2</sup>.** The ability of the myristoyl group to facilitate dynamic protein–protein and protein–membrane interactions (known as the myristoyl switch) makes it an essential feature of many signal transduction systems<sup>3</sup>. Thus pathogenic strategies that facilitate protein demyristoylation would markedly alter the signalling landscape of infected host cells. Here we describe an irreversible mechanism of protein demyristoylation catalysed by invasion plasmid antigen J (IpaJ), a previously uncharacterized *Shigella flexneri* type III effector protein with cysteine protease activity. A yeast genetic screen for IpaJ substrates identified ADP-ribosylation factor (ARF)1p and ARF2p, small molecular mass GTPases that regulate cargo transport through the Golgi apparatus<sup>4</sup>. Mass spectrometry showed that IpaJ cleaved the peptide bond between N-myristoylated glycine-2 and asparagine-3 of human ARF1, thereby providing a new mechanism for host secretory inhibition by a bacterial pathogen<sup>5,6</sup>. We further demonstrate that IpaJ cleaves an array of N-myristoylated proteins involved in cellular growth, signal transduction, autophagosome maturation and organelle function. Taken together, these findings show a previously unrecognized pathogenic mechanism for the site-specific elimination of N-myristoyl protein modification.

While seeking to understand how bacterial pathogens regulate the host secretory pathway<sup>7</sup>, we found that *S. flexneri* potently inhibited cargo transport through the Golgi apparatus (Supplementary Fig. 1). Further investigation showed that the Golgi was severely fragmented after *Shigella* infection (Fig. 1a). By contrast, neither *Listeria monocytogenes* nor *Salmonella typhimurium* disrupted the Golgi, suggesting that *Shigella* has evolved a specific, yet poorly understood, mechanism to regulate host organelle function.

*Shigella* infection requires translocation of more than 20 bacterial ‘effector’ proteins into host cells through the Mxi-SPA Type III Secretion System (T3SS)<sup>8</sup>. Deletion of *mxiD*, a component of the needle complex required for T3SS, eliminated Golgi destruction by *Shigella* infection (Fig. 1d). To identify the specific effector protein required for this activity, Golgi morphology was assessed after transient transfection of 20 effector genes (Supplementary Fig. 2a, b). Both IpaJ and VirA induced profound Golgi fragmentation (Fig. 1b) and inhibited hormone trafficking through the secretory pathway (Supplementary Fig. 2c). However, the other T3SS effector proteins had no discernible effect on Golgi structure or function (Supplementary Fig. 2a).

Comparison of cells infected with *Shigella ΔipaJ* and *ΔvirA* gene deletion strains showed that these mutants induced abnormal Golgi morphologies (Fig. 1c) and variable degrees of Golgi disruption (Fig. 1d), yet neither mutant alone fully attenuated Golgi destruction. By contrast, the Golgi remained intact and functional in cells infected with *ΔipaJ/virA* double mutant strain (Fig. 1c, d). The *Shigella ΔipaJ/virA* strain showed normal host cell invasion, intracellular replication and

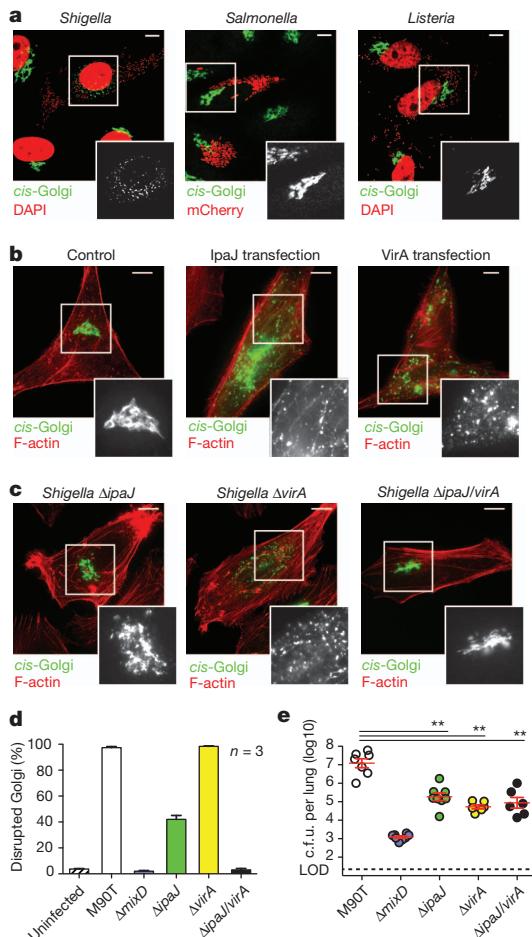
actin-based motility, suggesting Golgi disruption was specifically caused by these effector genes (Supplementary Fig. 3). Next, *in vivo* bacterial replication was evaluated using an established mouse model of mucosal infection<sup>9</sup>. The number of recoverable bacteria was sharply reduced (by 100-fold,  $P < 0.001$ ) by deleting either *ipaJ* or *virA* genes compared with wild-type *Shigella* (Fig. 1e). As expected, bacterial replication was further attenuated in the *Shigella ΔmixD* mutant lacking all T3SS function (Fig. 1e). We also found reductions in host inflammatory cytokines in mice infected with each *Shigella* mutant strain (data not shown). Taken together, these data demonstrate that two *Shigella* effector proteins, IpaJ and VirA, each harbour Golgi inhibitory activity, and are essential and play non-redundant roles for optimal *in vivo* virulence.

Although VirA was recently shown to inactivate Rab1 GTPase signalling pathways<sup>5</sup>, extra mechanisms are necessary for Golgi destruction during *Shigella* infection<sup>6</sup>. We therefore focused on IpaJ because its molecular mechanism of action is poorly understood<sup>10,11</sup>. BLAST database searches identified IpaJ homologues in numerous bacterial species, but this sequence-based alignment did not offer any clues to its function (Supplementary Fig. 4). In contrast, structural-based bioinformatics showed that IpaJ is distantly related to the cysteine peptidase C39-like family of domains of unknown function (annotated DUF3335) (Fig. 2a). C39-family members cleave leader peptides from bacteriocins with anti-microbial and quorum-sensing activities<sup>12</sup>. Although IpaJ is unlikely to function in these capacities, it harbours the catalytic cysteine (C), histidine (H) and aspartate (D) residues required for peptide bond hydrolysis (Fig. 2a). As predicted by this alignment, alanine substitutions at C64, H206 or D218 abolished the ability of IpaJ to disrupt host Golgi morphology in transfection studies (Fig. 2b). In addition, complementation of *Shigella ΔipaJ/virA* deletion strain with plasmid expression of IpaJ induced Golgi disruption phenotype whereas IpaJ C64A, H206A and D218A had no effect (Supplementary Fig. 5).

Previous studies have shown that IpaJ induces growth arrest phenotype in yeast<sup>13</sup>. A similar growth inhibitory activity was found with IpaJ, whereas yeast grew normally in the presence of IpaJ C64A, H206A or D218A catalytic mutants (Fig. 2c). Using a multicopy suppressor screen, we identified three distinct genomic loci that, when introduced on high-copy plasmids, suppressed IpaJ activity in yeast. Two loci encoded ARF1p and ARF2p GTPases, and the third loci encoded VPS15p, a phosphatidylinositol kinase required for yeast vacuole fusion<sup>14</sup> (Fig. 2d). Although these data suggest that IpaJ may regulate several host signalling pathways, we focused our initial efforts on ARF GTPases because these enzymes are master regulators of cargo trafficking through the Golgi apparatus<sup>4</sup>. Overexpression of either ARF1p or ARF2p rescued the yeast growth arrest phenotype, thereby defining ARF GTPases as potential substrates of IpaJ (Fig. 2e).

ARF1 functionally couples guanine-nucleotide exchange (GDP for GTP) and membrane binding (through N-myristylation) to the recruitment of COPI coat-complex to Golgi membranes as a key step

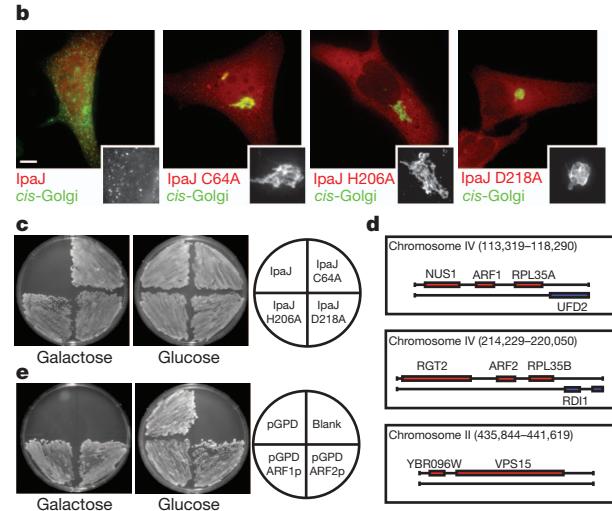
<sup>1</sup>Department of Microbiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-8816, USA. <sup>2</sup>Department of Pediatrics, Section of Pediatric Infectious Disease, Indiana University School of Medicine, 705 Riley Hospital Drive, ROC 4380, Indianapolis, Indiana 46202, USA. <sup>3</sup>Department of Pathology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-8816, USA. <sup>4</sup>Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 7017, Cincinnati, Ohio 45229-3039, USA.



**Figure 1 | Shigella IpaJ and VirA disrupt Golgi morphology.** **a**, Fluorescence microscopy HeLa cells infected with *Shigella* and *Listeria* visualized with 4',6-diamidino-2-phenylindole (DAPI) (pseudo-coloured red), and *Salmonella* carrying mCherry-expressing vector (red). The *cis*-Golgi (green) was detected by α-GM130 antibodies. Scale bar, 10 μm. **b**, Fluorescence microscopy of HeLa cells transfected with either IpaJ or VirA. The *cis*-Golgi (GM130, green) and F-actin (red) are shown. Scale bar, 10 μm. **c**, Fluorescence microscopy of HeLa cells infected with indicated *Shigella* mutant strains. Scale bar, 10 μm. **d**, Percentage of HeLa cells with disrupted Golgi morphology in 100 cells infected with either wild-type *Shigella* (M90T) or the indicated mutants. Error bars, means ± s.e.m. calculated from three independent experiments. **e**, Number of recoverable c.f.u. per mouse lung 24 h after intranasal inoculation with  $1 \times 10^6$  of wild-type M90T or each mutant *Shigella* strain as indicated. Limit of detection (LOD) for this assay is indicated by the dotted line. Error bars, geometrical means ± s.e.m. \*\*P < 0.001.

in vesicle fission<sup>4</sup>. Surprisingly, IpaJ had no effect on the guanine-nucleotide cycle or GTPase-dependent interactions of ARF1 *in vitro* (Supplementary Fig. 6). We therefore searched for IpaJ-induced post-translational modifications on ARF1 by ‘top-down’ mass spectrometry<sup>15,16</sup>. ARF1 bearing a carboxy (C)-terminal Strep affinity tag (ARF1-strep) was expressed in HEK293T cells and purified by Strep-Tactin chromatography. We observed an intact monoisotopic molecular mass of 26,895.95 Da and subsequent tandem mass spectrometry (MS/MS) resulted in 30 fragment ions that, together, provided an unambiguous assignment of *N*-myristoylated ARF1-strep protein (Fig. 3a and Supplementary Fig. 7a). Cellular expression of IpaJ resulted in a 266.35 Da decrease in molecular mass of ARF1-strep, which corresponds closely to the mass of the *N*-myristoyl group and the amino (N)-terminal glycine (267.22 Da) (Fig. 3b). Indeed, the MS/MS data support the idea that IpaJ cleaved the amide bond between glycine-2 and asparagine-3, liberating the *N*-myristoylated glycine from the ARF1 GTPase domain (Fig. 3d and Supplementary Fig. 7b). ARF1 cleavage

|                   | HHHHHHHHHHHH                   | EEEEEE EEEE                      |
|-------------------|--------------------------------|----------------------------------|
| IpaJ NP_085284    | 56-RKQMYDNSCGAAS1_LCAAKELG-79  | 204-GLHIVVLCRPDGS---YMDPAV-221   |
| C39L Q984Z8_RHIL0 | 23-YVQTPTFTCGAAC1_MMAMSALK-44  | 165-VPHIVVLAIGDDGHILIHDPWV-186   |
| C39L A6WZ5_OCHA4  | 23-YVQTPTFTCGAAC1_LMFAHHD-44   | 165-VPHIVVLAHGEDRHWIILIHDPWV-186 |
| C39L A0NWQ4_9RHOB | 23-YVQTPTFTCGAAC1_LMSFACLD-44  | 165-VPHIVVLAHGDADDRLHIIHDPMV-186 |
| C39L Q10032_9BACT | 24-YEQTDTFTCGAAC1_MMAGRDI-45   | 166-FPHIVVVTGFDEKVIVVFDFV-187    |
| C39L Q0EX59_9PROT | 24-YQQTCEFTCGAAC1_MMAMANFD-45  | 166-SPHIVVVSFGDRFRIVIHEPYV-187   |
| C39F YP_005847809 | 8-TKQHIDTDCCAACTATTSRQYG-33    | 92-LHYIVVIIHKTTKEVIVADPAK-113    |
| C1 AAB02650       | 150-KNQGSCGSCWAFAVSVTIEGII-171 | 290-VDHAVAAVGYPNPYILKNSWG-311    |

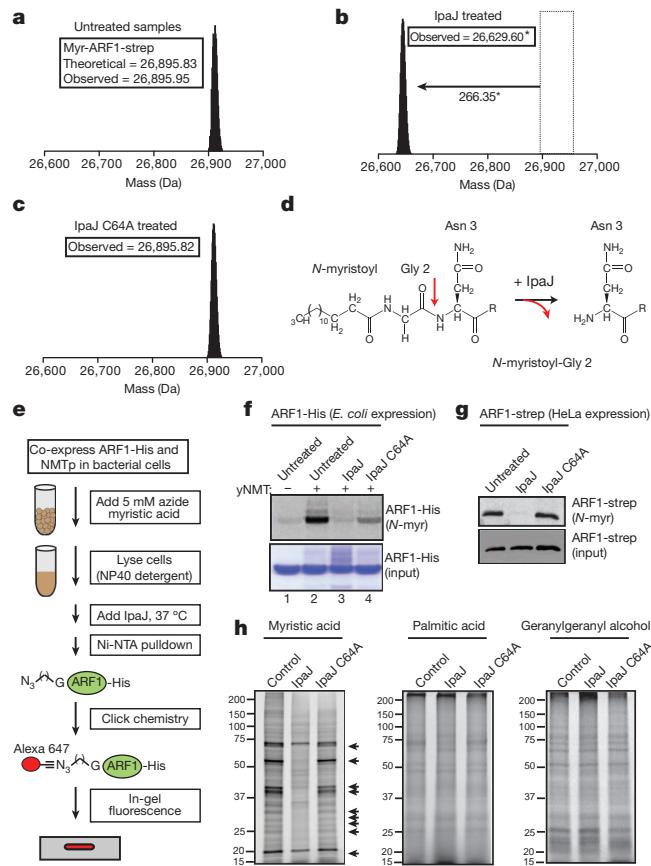


**Figure 2 | IpaJ belongs to the C39-like family of cysteine proteases and targets ARF-family GTPases.** **a**, Sequence-structure alignment of *S. flexneri* IpaJ with members of C39-peptidase-like family identified by HHpred. IpaJ possesses invariant catalytic triad residues Cys 64, His 206 and Asp 218 (red). The conserved Gln 58, which helps form the oxyanion hole found in many proteases, is shown in green. Invariant residues and core hydrophobic residues found in the C39-family are shown in blue and magenta, respectively. Characterized C39-family member (GCN5-acetyl transferase) and a C1-family member (Papain) are shown. **b**, Fluorescence microscopy of HeLa cells transfected with the indicated IpaJ mutants. The *cis*-Golgi (GM130, green) and F-actin (red) are shown. Scale bar, 10 μm. **c**, IpaJ or its catalytic mutants were expressed from a galactose-inducible promoter (pGal413 vector) and assayed for a growth arrest phenotype on galactose or glucose (control) carbon source. **d**, Illustration of the yeast genomic clones isolated from the IpaJ suppressor screen. ARF1p, ARF2p and VPS15p are highlighted. **e**, Yeast strain harbouring a galactose-inducible *IpaJ* gene were transformed with a multi-copy vector containing the indicated genes. Yeast were assayed for a growth arrest phenotype on galactose or glucose (control) carbon source.

was not observed in IpaJ C64A-treated samples, consistent with the function of IpaJ as cysteine protease (Fig. 3c and Supplementary Fig. 7c).

Next, we devised a protocol to reconstitute the ARF1 cleavage reaction *in vitro* (Fig. 3e). Because *Escherichia coli* does not express a system for *N*-myristylation, ARF1-His was co-expressed with yeast *N*-myristoyl transferase (NMTp) in the presence of exogenous azide-conjugated myristic acid. The lipid state of purified ARF1-His was visualized by fluorescently labelling the azido-myristoyl group by click chemistry<sup>17</sup> (Fig. 3f, lanes 1 and 2). Importantly, IpaJ cleaved *N*-myristoylated ARF1 *in vitro* whereas this reaction was less efficient with IpaJ C64A (Fig. 3f, lane 3 and 4). Similar results were also observed when IpaJ was incubated with ARF1-strep isolated from mammalian cellular lysates following a similar procedure (Fig. 3g).

Having established that IpaJ cleaves ARF1, we revisited our studies in yeast showing that IpaJ may target several host substrates including VPS15p, a lipidated kinase<sup>14</sup>. Although ARF1 and VPS15p are both *N*-myristoylated on their glycine-2 residue, they do not share any sequence or functional similarity. These observations suggested that IpaJ might cleave a broader range of substrates than initially suspected. To test this hypothesis, we labelled numerous of *N*-myristoylated proteins found in HEK293T cells and found that a large proportion of these proteins were demyristoylated in IpaJ-treated samples (Fig. 3h, arrows). The spectrum of *N*-myristoylated proteins was unaltered by IpaJ C64A mutant, thereby verifying the proteolytic-basis of this



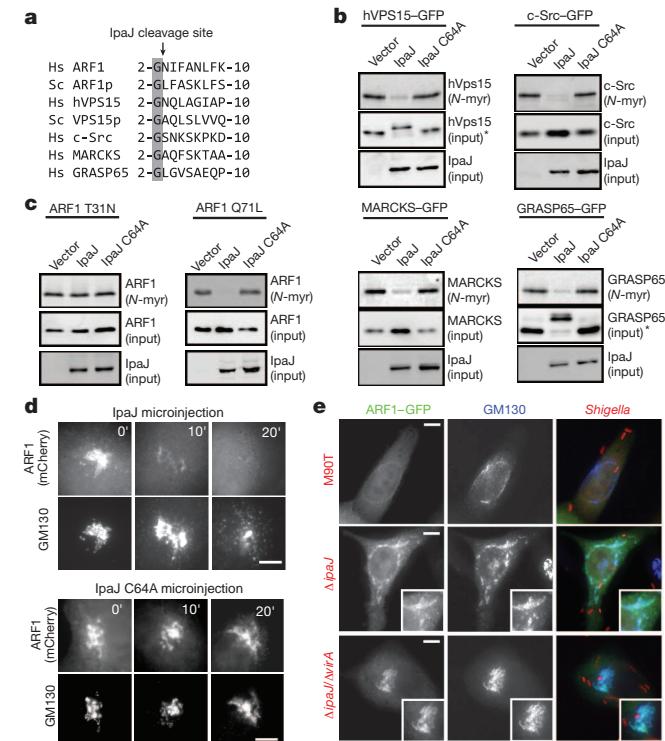
**Figure 3 | IpaJ cleaves the N-myristoylated glycine of lipidated substrates.** **a–c**, Mass spectra of purified ARF1-strep in untreated (a), IpaJ-treated (b) or IpaJ C64A-treated (c) cells, with the observed monoisotopic mass reported in daltons. \*Observed 1 Da shift in molecular mass is accounted for in MS/MS data (Supplementary Fig. 7). **d**, Structure of N-myristoylated ARF1 N terminus in untreated (left) and IpaJ-treated (right) samples. R group following Asn 3 denotes ARF1-strep protein residues 4–240. **e**, Reconstitution of ARF1-His N-myristylation (N-myrl) by NMTp in bacterial cells and IpaJ cleavage reaction *in vitro* (see Supplementary Information and Methods). **f**, In-gel fluorescence assay (top panel) visualizing Alexa Fluor 647-labelled myristoylated ARF1-His isolated from bacteria either not expressing (lane 1) or expressing (lane 2) NMTp. Bacterial cell lysates treated with recombinant IpaJ or IpaJ C64A as indicated. The expression levels of ARF1-His were determined by Coomassie blue stain. **g**, In-gel fluorescence assay (top panel) visualizing Alexa Fluor 647-labelled myristoylated ARF1-strep purified from HeLa cell lysates left untreated or incubated with IpaJ or its catalytic mutant as indicated. The expressed amounts of ARF1-strep are shown (bottom panel). **h**, In-gel fluorescence assay visualizing protein extracts isolated from HeLa cells incubated with azide myristic acid, azide palmitic acid or geranylgeranyl alcohol azide and subsequently labelled with Alexa Fluor 647 alkyne by click chemistry. Arrows indicate proteins that are proteolytically demyristoylated by IpaJ.

reaction. IpaJ had no effect on either palmitoylation or geranylgeranylation, indicating that the protease specifically cleaves proteins modified by the myristoyl group (Fig. 3h).

N-myristoylated proteins share a conserved glycine but their sequences are otherwise highly variable (Fig. 4a). To determine how IpaJ recognizes diverse substrates, we first generated chimaeric enhanced green fluorescent protein (EGFP) fusion proteins that expressed 10 amino-acid leader peptides derived from c-Src, MARCKS, hVPS15 and GRASP65, N-myristoylated proteins involved in cellular growth, signal transduction, autophagosome maturation and organelle function, respectively (Fig. 4a)<sup>18–21</sup>. IpaJ cleaved each of these lipid-modified peptides in a reaction requiring its cysteine protease activity (Fig. 4b). The ability of IpaJ to cleave unrelated peptides suggested that the protease might recognize the lipidated glycine-2 common to all of its substrates. We found that soluble myristic acid was unable to inhibit

ARF1 proteolysis, suggesting that IpaJ does not simply recognize long hydrocarbon chains (Supplementary Fig. 8). However, protein lipidation was essential for proteolysis because IpaJ was unable to cleave non-myristoylated ARF1 *in vitro* (Supplementary Fig. 9). When taken together, these data indicate that IpaJ specifically recognizes the conserved glycine-2 of substrates in context of its amide linkage to N-myristic acid.

N-myristylation is unique among lipid modifications in that it undergoes a dynamic conformational switch in response to upstream signalling events (termed the ‘myristoyl switch’)<sup>3</sup>. For example, the myristoyl group is sequestered in GDP-inactive ARF1 and liberated for membrane binding and signalling by GTP exchange (Supplementary Fig. 10)<sup>4,22–24</sup>. We found that IpaJ was unable to cleave ARF1 T31N (a GDP-bound mutant form), but efficiently cleaved ARF1 Q71L (a mutant locked in the GTP-bound active state) in cells (Fig. 4c). These findings suggest IpaJ recognizes substrates as they cycle between activity-dependent conformational states. Consistent with this interpretation, microinjection of purified IpaJ induced the release of activated ARF1 from Golgi membranes just before organelle fragmentation (Fig. 4d). ARF1 was also displaced from Golgi membranes by wild-type *Shigella* infection but not by  $\Delta$ IpaJ single mutant (note that the



**Figure 4 | Conformational-dependent cleavage of lipidated substrates by IpaJ.** **a**, Multiple-sequence alignment of N-myristoylated glycine and downstream sequences from the indicated proteins (Hs, *Homo sapiens*; Sc, *S. cerevisiae*). **b**, In-gel fluorescence assay showing myristoylated peptides of the indicated proteins in cells expressing IpaJ or catalytic mutant. The peptide-eGFP and IpaJ inputs are indicated. \*Demiyristoylation of GRASP65 and hVPS15 resulted in slower mobility of the resulting peptide, potentially owing to reduced mobility in SDS-polyacrylamide gel electrophoresis caused by proteolytic reaction. **c**, In-gel fluorescence assay showing myristoylated ARF1 T31N mutant (a GDP-locked mutant) or ARF1 Q71L (a GTP-locked mutant) after in cells co-expressing IpaJ or catalytic mutants as indicated. Equal loading of ARF1 (middle panel) and IpaJ (lower panel) was determined by western blot analysis. **d**, Fluorescence microscopy showing ARF1-mCherry and the Golgi apparatus (GM130) after cellular microinjection of recombinant IpaJ or IpaJ C64A, over the indicated period. Scale bar, 10  $\mu$ m. **e**, Fluorescence microscopy showing ARF1-eGFP (green) and Golgi membranes (blue, GM130) after infection with *Shigella* M90T and the indicated mutants (red; mCherry-expressing bacteria). Scale bar, 10  $\mu$ m.

Golgi is fragmented by VirA in this infection) or  $\Delta ipaJ/\Delta virA$  double mutant strains (Fig. 4e). Extra cell-based experiments support the notion that GTP-active ARF1 is a physiological target of IpaJ during *Shigella* infection (Supplementary Figs 11–13), thereby defining a new mechanism for inhibition of cytokine secretion and immune detection by a bacterial pathogen.

Here we have determined that IpaJ is a cysteine protease that cleaves the N-myristoylated glycine from ARF1 and potentially numerous host cellular substrates. This discovery adds significantly to the small but increasing number of bacterial proteases that cleave lipidated proteins as a mechanism of microbial pathogenesis<sup>25,26</sup>. Importantly, the ability of IpaJ to eliminate the N-myristoyl modification of unrelated proteins may allow *Shigella* to exploit signalling pathways in a variety of host cellular contexts<sup>27</sup>. Although extra studies will be needed to unravel the hierarchy of these events, our discoveries provide a unique framework from which to explore the functional consequences of protein N-myristylation in human health and disease.

## METHODS SUMMARY

All bacterial effector genes were amplified by PCR from *S. flexneri* M90T genomic DNA and cloned into the relevant expression vectors. *S. flexneri* M90T knockout strains were generated using a λ red recombinase-mediated recombination system<sup>28</sup>. For the mucosal model of infection, *S. flexneri* strains were administered dropwise ( $10^6$  colony-forming units (c.f.u.)) into the external nares of mice, and bacterial c.f.u. were enumerated by plating serial dilutions of lung homogenates as described previously<sup>9</sup>. For structural bioinformatics, HHpred was used to detect known Pfam domains with distant structural homology relationships, and PROMALS3D<sup>29</sup> was used to produce a multiple-sequence alignment to C39-like family members. We performed the yeast suppressor screen by integrating a single copy of the *ipaJ* under the Gal1 promoter into Y7092 yeast strain genome. A *Saccharomyces cerevisiae* genomic library (Yep13 vector, American Type Culture Collection) was transformed into yeast and positive library clones were mapped by DNA sequencing. To identify post-translational modifications of IpaJ, liquid chromatography (LC)–MS/MS was performed as previously described<sup>30</sup>. To detect N-myristylation of proteins, cells were metabolically labelled with myristic acid-azide<sup>17</sup>, and purified N-myristoylated proteins were labelled with Alexa Fluor 647 Alkyne using Click-iT Reaction Buffer (Invitrogen).

**Full Methods** and any associated references are available in the online version of the paper.

Received 21 September 2012; accepted 12 February 2013.

Published online 27 March 2013.

- Farazi, T. A., Waksman, G. & Gordon, J. I. The biology and enzymology of protein N-myristylation. *J. Biol. Chem.* **276**, 39501–39504 (2001).
- Maurer-Stroh, S. et al. MYRbase: analysis of genome-wide glycine myristylation enlarges the functional spectrum of eukaryotic myristoylated proteins. *Genome Biol.* **5**, R21 (2004).
- McLaughlin, S. & Aderem, A. The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions. *Trends Biochem. Sci.* **20**, 272–276 (1995).
- Kahn, R. A. Toward a model for Arf GTPases as regulators of traffic at the Golgi. *FEBS Lett.* **583**, 3872–3879 (2009).
- Dong, N. et al. Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell* **150**, 1029–1041 (2012).
- Mounier, J. et al. *Shigella* effector IpaB-induced cholesterol relocation disrupts the Golgi complex and recycling network to inhibit host cell secretion. *Cell Host Microbe* **12**, 381–389 (2012).
- Selyunin, A. S. et al. The assembly of a GTPase-kinase signalling complex by a bacterial catalytic scaffold. *Nature* **469**, 107–111 (2011).
- Ashida, H. et al. *Shigella* are versatile mucosal pathogens that circumvent the host innate immune system. *Curr. Opin. Immunol.* **23**, 448–455 (2011).
- Way, S. S., Borczuk, A. C., Dominitz, R. & Goldberg, M. B. An essential role for gamma interferon in innate resistance to *Shigella flexneri* infection. *Infect. Immun.* **66**, 1342–1348 (1998).
- Buyssse, J. M., Dunyak, D. S., Hartman, A. B. & Venkatesan, M. M. Identification and molecular characterization of a 27 kDa *Shigella flexneri* invasion plasmid antigen, IpaJ. *Microb. Pathog.* **23**, 357–369 (1997).
- Liu, M. et al. Sequence analysis and characterization of plasmid pSFD10 from *Salmonella choleraesuis*. *Plasmid* **48**, 59–63 (2002).

- Michiels, J., Dirix, G., Vanderleyden, J. & Xi, C. Processing and export of peptide pheromones and bacteriocins in Gram-negative bacteria. *Trends Microbiol.* **9**, 164–168 (2001).
- Slagowski, N. L., Kramer, R. W., Morrison, M. F., LaBaer, J. & Lesser, C. F. A functional genomic yeast screen to identify pathogenic bacterial proteins. *PLoS Pathog.* **4**, e9 (2008).
- Herman, P. K., Stack, J. H., DeModena, J. A. & Emr, S. D. A novel protein kinase homolog essential for protein sorting to the yeast lysosome-like vacuole. *Cell* **64**, 425–437 (1991).
- Sze, S. K., Ge, Y., Oh, H. & McLafferty, F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. USA* **99**, 1774–1779 (2002).
- Kelleher, N. L. Top-down proteomics. *Anal. Chem.* **76**, 197A–203A (2004).
- Charron, G. et al. Robust fluorescent detection of protein fatty-acylation with chemical reporters. *J. Am. Chem. Soc.* **131**, 4967–4975 (2009).
- Barr, F. A., Puype, M., Vandekerckhove, J. & Warren, G. GRASP65, a protein involved in the stacking of Golgi cisternae. *Cell* **91**, 253–262 (1997).
- Kamps, M. P., Buss, J. E. & Sefton, B. M. Mutation of NH<sub>2</sub>-terminal glycine of p60src prevents both myristylation and morphological transformation. *Proc. Natl. Acad. Sci. USA* **82**, 4625–4628 (1985).
- Stumpo, D. J., Graff, J. M., Albert, K. A., Greengard, P. & Blackshear, P. J. Molecular cloning, characterization, and expression of a cDNA encoding the “80- to 87-kDa” myristoylated alanine-rich C kinase substrate: a major cellular substrate for protein kinase C. *Proc. Natl. Acad. Sci. USA* **86**, 4012–4016 (1989).
- Sun, Q. et al. Identification of Barkor as a mammalian autophagy-specific factor for Beclin 1 and class III phosphatidylinositol 3-kinase. *Proc. Natl. Acad. Sci. USA* **105**, 19211–19216 (2008).
- Franco, M., Chardin, P., Chabre, M. & Paris, S. Myristylation-facilitated binding of the G protein ARF1GDP to membrane phospholipids is required for its activation by a soluble nucleotide exchange factor. *J. Biol. Chem.* **271**, 1573–1578 (1996).
- Liu, Y., Kahn, R. A. & Prestegard, J. H. Structure and membrane interaction of myristoylated ARF1. *Structure* **17**, 79–87 (2009).
- Liu, Y., Kahn, R. A. & Prestegard, J. H. Dynamic structure of membrane-anchored Arf\*GTP. *Nature Struct. Mol. Biol.* **17**, 876–881 (2010).
- Shao, F., Merritt, P. M., Bao, Z., Innes, R. W. & Dixon, J. E. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* **109**, 575–588 (2002).
- Choy, A. et al. The *Legionella* effector RavZ inhibits host autophagy through irreversible Atg8 deconjugation. *Science* **338**, 1072–1076 (2012).
- Philapon, A. & Sansonetti, P. J. Shigellosis: innate mechanisms of inflammatory destruction of the intestinal epithelium, adaptive immune response, and vaccine development. *Crit. Rev. Immunol.* **23**, 371–401 (2003).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
- Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
- Roth, M. J. et al. Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry. *Anal. Chem.* **83**, 9586–9592 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank our colleagues at University of Texas Southwestern Medical Center, specifically K. Orth, Y. M. Chook and J. Seeman, for discussions in preparing this manuscript. We are particularly indebted to L. Kinch, N. Grishin, D. Mitchel, X. Guo, D. Trudgian and S. Perelman for their contributions. We acknowledge the services of the University of Texas Southwestern Medical Center proteomics core, supported by a Cancer Prevention and Research Institute of Texas grant RP120613. A.S.S. was supported by the Howard Hughes Medical Institute International Student Research fellowship. B.A.W. was supported by a National Institutes of Health training grant (NIHAIID; 5T32AI007520) and S.S.W. by the Burroughs Wellcome Fund. This work was supported by grants from the National Institutes of Health (NIHAIID; R01AI083359 and NIGMS; R01GM100486), the Welch Foundation (I-1704) and the Burroughs Wellcome Fund to N.M.A.

**Author Contributions** N.B. and N.M.A. conceived the general ideas for this work. N.B. and N.M.A. planned, performed and interpreted experiments. T.G.F. and A.S.S. generated bacterial knockout strains and performed GTPase assays, respectively. J.M.E. and S.S.W. performed mouse infection studies. B.A.W. performed the bioinformatics studies on IpaJ. D.A.P. and S.M.P. performed top-down mass spectrometry. N.M.A. and N.B. wrote the manuscript and all authors provided editorial input.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.M.A. (neal.alto@utsouthwestern.edu).

## METHODS

**Plasmids and cloning.** Bacterial effector genes were amplified by PCR from *S. flexneri* M90T and cloned into pEntr/D, generating a Gateway<sup>TM</sup> compatible entry clone (Invitrogen) according to the manufacturer's recommendations. Twenty *S. flexneri* effectors were cloned into pEntr/D vector: IpaA, AAK18443.1; IpaB, AAK18446; IpaH1.4, AAK18594.1; IpaH2.5, AAK18367; IpaH4.5, AAK18395; IpaH7.8, AAK18394.1; IpaH9.8, AAK18544; IpaJ, AAK18440; IpgD, AAK18452; IcsB, AAK18450; OspB, AAL72323.1; OspC1, AAL72322.1; OspC2, AAW64906; OspD1, AAW64782; OspE1, AAW64916; OspE2, AAW64805; OspF, AAW64770; OspG, NP\_085391; VirA, AAK18501. For mammalian expression, the effectors were then recombined into modified pcDNA3.1 vector carrying N-terminal eGFP or mCherry in frame with the Gateway cassette (Gateway<sup>TM</sup> vector conversion system, Invitrogen). For bacterial expression, a 50-amino-acid N-terminal deletion of IpaJ (IpaJΔ50; residues 51–259) was cloned in frame into pGEX-4T1 (GST-tag) (Amersham), yielding intact catalytic domain of IpaJ C-terminally fused to GST-tag. N-terminal deletion of ARF1 (ARF1Δ17; residues 18–181) was subcloned by PCR into pGEX-4T1 vector. Human GGA1 (GAT domain, residues 76–215) was amplified and cloned by PCR in frame into pET28b-MBP-His vector. For complementation of *S. flexneri* knockout strains, *ipaJ* was amplified and cloned by PCR into pBad/Myc-His vector to be expressed under arabinose-inducible promoter. mCherry was expressed from the *rpsD* promoter in pDP151 vector in *Shigella* strains. For galactose-induced yeast expression, *ipaJ* was recombined from pEntr/D into Gateway<sup>TM</sup>-compatible pYes-Dest52 vector (Invitrogen) for expression under galactose-inducible promoter. Yeast *arf1* and *arf2* (SGD: S000002351, SGD: S000002296) were amplified by PCR from YEP13 genomic clones and cloned into p415 vector for yeast expression from a constitutively active promoter (GAPDH). Human full-length *arf1* was cloned in frame into modified pcDNA3.1 vector carrying C-terminal mCherry or eGFP. Human GRASP65 was cloned in frame into pEGFP-N2 vector. For expression of Strep-tagged full-length human ARF1 in mammalian cells, the *arf1* complementary DNA (cDNA) (Missouri S&T cDNA Resource Center; ARF0100000) was first cloned into pEntr/D entry vector. The gene was then recombined into modified pcDNA4T/O vector carrying Gateway<sup>TM</sup> cassette and two repeats of C-terminal Strep-tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys, IBA GmbH). For expression of leader peptides bearing myristylation signal sequences derived from c-Src, MARCKs, hVPS15 and GRASP65, cDNA sequences encoding 10 N-terminal amino acids (see Fig. 4a) were cloned into modified pcDNA4T/O vector carrying C-terminal GFP and two repeats of Strep-tag. All site-directed mutations were generated using the QuickChange Site-Directed Mutagenesis Kit (Stratagene). More primer and plasmid information is available upon request.

**Bacterial infection of cultured cells.** For *Shigella* infections, HeLa cells or HeLa cells stably expressing a GFP tag attached to the N-terminal Golgi retention signal of *N*-acetylglucosaminyltransferase I<sup>31</sup> (NAGFP HeLa) were seeded onto coverslips in six-well dishes the day before infection. Sodium butyrate at 5 μM was added to NAGFP cells after seeding to stimulate production of Golgi-localized GFP marker<sup>31</sup>. *Shigella* strains were inoculated from frozen stocks and grown overnight at 30 °C in brain-heart infusion media (BHI) (Difco<sup>TM</sup>, BD Biosciences). Bacteria were then back-diluted 1:50 and incubated at 37 °C until reaching  $D_{600} \sim 0.5\text{--}0.6$ . Bacteria were then washed in 1× phosphate buffer saline (PBS) and incubated at 37 °C for 15 min in 0.003% Congo red. Twenty microlitres of bacterial suspension were then added to each well of a six-well dish of semi-confluent HeLa or NAGFP cells and centrifuged for 10 min at room temperature (1000g) to facilitate bacterial adherence. The plates were then incubated for 90 min at 37 °C, 5% CO<sub>2</sub>. The media was removed and the wells were washed three times with gentamicin (150 μg ml<sup>-1</sup>) followed by three washes with sterile PBS. Fresh antibiotic-free DMEM was added to each well after that and cells were incubated for an extra 2–4 h (37 °C, 5% CO<sub>2</sub>). After incubation, the slides were fixed in 3.7% formaldehyde and prepared for microscopy. Golgi and F-actin were visualized with anti-GM130 (BD Transduction Laboratories) and rhodamine-phalloidin (Molecular Probes) respectively. For infections with *S. flexneri* knockout strains complimented with IpaJ (wild type and C/H/D mutants), overnight bacterial cultures were diluted 1:50 in BHI containing 2% arabinose and incubated until reaching  $D_{600} \sim 0.5\text{--}0.6$  as described above. Infections were performed in low-glucose media containing 2% arabinose to allow expression of IpaJ from arabinose-inducible promoter. For infections with *S. flexneri* strains expressing mCherry, ampicillin was added both to BHI and DMEM culture media.

A protocol for *Salmonella* infection of HeLa cells was adapted from previous studies<sup>32</sup>. *S. typhimurium* was grown overnight at 37 °C in a glass flask with shaking then subcultured (1:30) and grown for 3 h. One millilitre was pelleted, suspended in 1× PBS and added dropwise at various concentrations to semi-confluent HeLa cells in low-glucose DMEM+10% FBS. Cells were incubated (37 °C, 5% CO<sub>2</sub>) for 10 min, washed three times with PBS and incubated for an extra 15 min (37 °C, 5% CO<sub>2</sub>) in fresh low-glucose DMEM + 10% FBS. Cells were

washed again with PBS and incubated in DMEM + 10% FBS + 100 mg ml<sup>-1</sup> gentamicin before fixing with 3.7% formaldehyde at various time points. Cellular phenotypes were visualized as described.

*L. monocytogenes* was grown and prepared for HeLa infection as described<sup>33</sup>. Various concentrations of bacteria were added to semi-confluent HeLa cells in DMEM + 10% FBS from an overnight culture of *Listeria* grown in BHI media. Cells were centrifuged (250g, room temperature, 5 min) and incubated (37 °C, 5% CO<sub>2</sub>) for 10 min, then washed three times with 1× PBS. Fresh media containing 10 mg ml<sup>-1</sup> gentamicin were added and cells incubated for an extra 4–6 h before fixation and visualization.

**Gene disruption in *Shigella*.** The *virA*, *ipaJ* and *mxiD* genes were individually disrupted using the λ red recombinase-mediated recombination system<sup>28</sup>. Briefly, a kanamycin resistance cassette flanked with 50 base pairs homologous to the gene of interest (*virA*, *ipaJ* or *mxiD*) was amplified from plasmid DNA (pKD3) (primer sequences are available upon request). PCR products were electroporated into *S. flexneri* strain M90T carrying the red recombinase plasmid pKD46. Transformants were selected by growth on Luria broth agar plates containing kanamycin (50 μg ml<sup>-1</sup>) and simultaneously cured of pKD46 by growth at 42 °C overnight. The kanamycin resistance gene was eliminated through the introduction of the pCP20 helper plasmid that contained the FLP recombinase. Subsequent curing of pCP20 was done by growing strains at 42 °C for 5 h. Disruption of the *virA* and *ipaJ* genes was confirmed through DNA sequencing of the respective genetic loci. To generate the double strain (*Shigella* Δ*virA*/Δ*ipaJ*), the *ipaJ* locus was disrupted from a *Shigella* Δ*virA* strain following the protocol described above.

**Shigella infection studies in mice.** Female C57BL/6 mice were purchased from the National Cancer Institute and used between 6 and 8 weeks of age. For infection, *S. flexneri* M90T or mutants were grown and back-diluted from the stationary to early log phase growth ( $D_{600} \sim 0.1$ ) in BHI media (BD Biosciences) at 37 °C, washed and diluted in sterile saline, and administered dropwise (10<sup>6</sup> c.f.u. in 30 μl volume) into the external nares of mice anaesthetized with ketamine/xylazine as described<sup>9</sup>. At the indicated time points after infection, mice were euthanized, and the recoverable bacterial c.f.u. enumerated by plating serial dilutions of the lung homogenate on BHI plates. Statistical significance was determined by a one-way analysis of variance ( $P < 0.01$ ) Tukey post hoc test. The amounts of the TNF-α and IL-6 in the serum and lung homogenate were enumerated using commercially available enzyme-linked immunosorbent assay (ELISA) reagents (R&D Systems). All experiments were performed under University of Minnesota or Cincinnati Children's Hospital Institutional Animal Care and Use Committee approved protocols (S.S.W.).

**hGH trafficking assay.** As previously described<sup>7,34</sup>, HeLa cells (50% confluence) were transfected with 1 μg of 4X FKBP-hGH (Ariad Pharmaceutical; source of material, D. Bernstein) and either 0.5 μg eGFP-IpaJ, eGFP-VirA or pEGFP control plasmid with Fugene6 (Roche). Sixteen hours later, the medium was replaced with medium containing AP21998 (final concentration 2 mM) or vehicle control. AP21998 was incubated with the cells for 2 h before the supernatant was collected. The supernatant was then diluted 100-fold and compared against an hGH standard curve (12.5–400 pg ml<sup>-1</sup>) for the quantification of hGH released using an hGH enzyme-linked ELISA (Roche). For no drug controls, 100% ethanol (2 μl) was incubated with the cells for 2 h.

**IpaJ bioinformatics.** HHpred was used to detect known Pfam domains (profile database search used: pfam version 25.0) with distant structural homology relationships to full-length IpaJ (GenInfo Identifier (GI): 12329066) with default parameter settings<sup>35</sup>. The DUF3335 (C39-like peptidase family) gave a probability score of 90.0. One hundred proteins in the Pfam database contain this domain. PROMALS3D was used to produce a multiple-sequence alignment with the eight IpaJ and 100 DUF3335 family members to identify invariant catalytic residues and confirm conserved secondary and hydrophobicity patterns<sup>29</sup>.

**Recombinant protein expression and purification.** Recombinant proteins (IpaJΔ50, ARF1ΔAN17, GGA1<sub>176–215</sub>) were expressed in BL21-DE3 *E. coli* strains by induction with 0.4 mM IPTG for 16 h at 18 °C. Pellets were lysed with His buffer (100 mM HEPES, pH 7.5, 300 mM NaCl) or GST buffer (Tris buffer saline (TBS); 50 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT) supplemented with protease cocktail (Roche). Proteins were purified with nickel agarose (Qiagen) or glutathione Sepharose (Amersham Biosciences) following the manufacturer's instructions. Eluted proteins were buffer exchanged into TBS using concentration centrifugal columns (Millipore); glycerol was added to 15% and the proteins were then stored at –80 °C.

**Cell transfections, microinjections and fluorescence microscopy.** HEK293A, HeLa and NAGFP HeLa (see above) cells were transfected using FuGene6 (Roche) and incubated for 16–18 h. For expression of Strep-tagged proteins, HEK293T cells were transfected using calcium phosphate and incubated for 18–24 h. Equal amounts of DNA were used for co-transfection. Cells were then lysed (lysis buffer: 20 mM Tris HCl pH 7.5, 1.5 mM MgCl<sub>2</sub>, 350 mM NaCl, 0.5% NP-40, 5% glycerol)

and sonicated for 5 s. The lysate was then clarified by centrifugation (15,000g, 10 min) and applied to Strep-Tactin Superflow Plus resin (Qiagen). After incubation (90 min, 4 °C) the column was washed three times (washing buffer: 20 mM Tris HCl pH 7.5, 1.5 mM MgCl<sub>2</sub>, 150 mM NaCl, 0.2% NP-40, 5% glycerol) and Strep-tagged proteins were eluted with 2.5 mM desthiobiotin (100 mM Tris·HCl, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin, pH 8.0). Microinjections of IpaJΔ50 were performed using a semiautomatic InjectMan NI2 micromanipulator (Eppendorf). Recombinant proteins were diluted in 1× TBS with Cascade Blue (Invitrogen) fluorescent dyes until the final concentration of the protein of 1 mg ml<sup>-1</sup> (21 μM). Cellular concentration of microinjected protein was estimated as 1 μM. For IpaJ C64A and ARF1 co-localization studies, HeLa cells were treated either with Brefeldin A (2.5 μg ml<sup>-1</sup>) for 20 min or with nocodazole (10 μg ml<sup>-1</sup>) for 1 h before fixation. Immunofluorescence images in all experiments were acquired with Zeiss AxioVert 200 fluorescence microscope. F-actin was visualized by rhodamine- or Alexa Fluor 350-conjugated phalloidin (Molecular Probes), Golgi was detected by GM130 antibodies (BD Biosciences) and secondary anti-mouse IgG antibodies (Thermo Scientific). NAGFP HeLa cells stimulated for production of GFP marker were additionally stained with anti-GFP antibodies (Clontech) and fluorescein-conjugated anti-rabbit IgG secondary antibodies (Thermo Scientific) to enhance the fluorescent signal.

**Galactose-induced yeast growth inhibition and yeast multicopy suppressor screen.** Yeast InvSc1 strain was transformed with pYES-dest52 (Invitrogen) vector carrying *ipaJ* gene (wild type or C/H/D catalytic mutant) under the GAL1 inducible promoter. Yeast was streaked onto Yc-U agar media containing glucose or galactose and raffinose as carbon source. Yeast was cultured for at 30 °C for 3–5 days and the plates were visually inspected for growth. For the yeast multicopy suppressor screen, the *ipaJ* gene was stably integrated into Y7092 yeast strain genome (MAT $\alpha$  *can1Δ*:STE2pr-Sp<sub>5</sub>*his5* *his3Δ1leu2Δura3Δ0met15Δ0 lyp1Δ*) as previously described<sup>36</sup>. The resulting strain Y7092-IpaJ (MAT $\alpha$  *can1Δ*:STE2pr-Sp<sub>5</sub>*his5* *his3Δ1leu2Δura3Δ0met15Δ0 lyp1Δ trp1Δ::GAL1-*ipaJ*-URA3) carried a single copy of the *ipaJ* gene under control of GAL1 galactose-inducible promoter. Y7092-IpaJ cells were then transformed with the *S. cerevisiae* genomic library in the Yep13 vector (ATCC) and transformants were selected for survival on Yc-UL media containing 2% galactose and 1% raffinose as carbon sources. Positive clones were isolated and library vectors were sequenced. For the complementation assay, PCR-amplified yeast *arf1* and *arf2* were expressed in yeast under GAPDH promoter (see above).*

**Top-down mass spectrometry.** Strep-tagged ARF1 protein was co-transfected with IpaJ or control vector. ARF1-strep was purified with a Strep-Tactin Superflow Plus column and eluted with 2.5 mM desthiobiotin buffer (see above). Purified ARF1-strep was used for LC-MS/MS analysis.

LC-MS/MS was generally as previously described<sup>30</sup>. Optima-grade solvents and acids (Thermo Scientific) were used. Reverse-phase liquid chromatography (RPLC) capillary columns were packed in-house to a length of 15 cm with 5 μm diameter C18 Poroshell-300 resin (Agilent Technologies) in 75 μm inner diameter × 360 μm outer diameter Picofrit columns (New Objective) with 15 μm inner diameter integrated micro-electrospray tips. A capillary column heater (Analytical Sales & Services) was used to maintain the column temperature at 60 °C during analysis. RPLC mobile phase A included 0.025% TFA, 0.3% formic acid and 5% acetonitrile in water. RPLC mobile phase B included 0.025% TFA, 0.3% formic acid and 20% isopropanol in acetonitrile. The elution gradient was 0% B at 0–3 min, 45% B at 3.01 min and 45–60% B from 3.01 to 23 min. Flow was regulated by an 1100-nano-LC (Agilent) at a flow rate of 0.5 μl min<sup>-1</sup>.

Before analysis, the protein was de-salted using a C4 ZipTip (Millipore) as described previously and suspended in 0.2% formic acid in water. Approximately 1600 fmol of protein was loaded onto the capillary column per injection. Analysis was on an LTQ Orbitrap XLT (Thermo Scientific) in full MS mode at 60,000 Fourier-transform mass spectrometry resolution, two microscans, maximum ion accumulation time of 500 ms, scan range from 800 to 2000 *m/z*, source-induced dissociation = 25 V, tube lens 100 V, capillary voltage 50 V, capillary temperature 275 °C. MS/MS data were collected with the source-induced dissociation = 60 V, tube lens 100 V, capillary voltage 50 V, capillary temperature 375 °C. The [M + 6H]<sup>6+</sup> charge state of ubiquitin was used to tune the mass spectrometer. The 'Xtract' function in the XcaliburTM (Thermo Scientific) data system with a signal/noise of 3 was used to extract intact protein and fragment masses from the raw spectral data. All fragmentation data presented were the summation of scans across the entire protein elution profile. MS/MS data sets were interrogated with Prosight PC 2.0TM (Thermo Scientific). Analysis of the MS/MS data against the known ARF1 sequences without the correct modification

state led to few or no b-ion matches. Replicate analysis was performed on each sample and assigned fragment ions were manually validated in raw data.

**Labelling N-myristoylated proteins by click chemistry and *in vitro* cleavage reaction.** *E. coli* BL-21 cells expressing ARF1-His with or without yeast NMTp were grown overnight at 37 °C in LB media. Cultures were diluted 1:50 and incubated at 37 °C until reaching *D*<sub>600</sub> ~ 0.5–0.6. Myristic acid and azide myristic acid were added at the concentrations 50 μM and 5 μM respectively, and cells were additionally incubated for 30 min. Protein expression was then induced with 0.4 mM IPTG and cells were further incubated at 37 °C for 3 h. We estimated that N-myristoylated ARF1 is bound in a GDP:GTP ratio of 1:1 as assessed by GGA-binding assays. Expression of IpaJ or IpaJ C64A mutant was performed by the same method without adding myristic acid. ARF1- and IpaJ-expressing cells were lysed by sonication for 1 mn separately (lysis buffer: 20 mM Tris HCl pH 7.5, 1.5 mM MgCl<sub>2</sub>, 350 mM NaCl, 0.5% NP-40, 5% glycerol) and the lysates were mixed together and incubated for 30 min at 37 °C. ARF1 was then purified using nickel agarose (Qiagen). N-myristoylated proteins were labelled with Alex Fluor 647 Alkyne using Click-iT Reaction Buffer Kit (Invitrogen) according to instructions with modifications. Specifically, click labelling was performed on the proteins still bound to the column. The column was then washed and proteins were eluted with 500 mM imidazole/100 mM HEPES containing 1% SDS. The N-myristylation status was analysed by in-gel fluorescence and the equal protein load was confirmed by Coomassie stain. For cleavage inhibition, free myristic acid or vehicle control (DMSO) was added to bacterial lysate containing ARF1 before adding IpaJ. After cleavage reaction, ARF1 was then labelled and purified as described above. To test the cleavage of non-myristoylated ARF1, a similar experimental setup was used with the following modifications: *E. coli* BL-21 cells expressing ARF1-His (in the absence of NMTp or exogenous myristic acid) were purified with Ni-NTA agarose, eluted with 500 mM imidazole/100 mM HEPES and processed for mass spectrometry.

For metabolic labelling in mammalian cells, myristic acid-azide (Invitrogen) was added to the cell cultures 6 h after transfection at final concentration 10 μM and incubation was continued overnight. The next day, Strep-tagged proteins were purified using Strep-Tactin SuperFlow Plus columns (see above). N-myristoylated proteins were labelled with Alex Fluor 647 Alkyne as described above. The column was then washed and proteins were eluted with 2.5 mM desthiobiotin elution buffer. Purified proteins were separated on SDS-polyacrylamide gel electrophoresis gel and transferred onto nitrocellulose paper (Biorad) for probing with STREPtactin-HRP (Biorad). Fluorescence was analysed both in SDS-gel and on nitrocellulose paper (collectively 'in-gel fluorescence'). Expression of IpaJ-GFP (wild type or catalytic mutants) was confirmed by probing cell lysates with anti-GFP antibodies (Clontech) and secondary HRP-anti-rabbit IgG antibodies (Invitrogen). Expression of GRASP65- and hVPS15-GFP chimaeric proteins was probed in cell lysate, before protein purification. For *in vitro* cleavage reaction, ARF1-strep transfected cells labelled with myristic acid-azide were lysed and 3 μg of recombinant IpaJ Δ50 (wild type or C64A catalytic mutant) were added into 20 μl of the lysate. After incubation at 37 °C for 30 min, myristoylated ARF1-strep was labelled and analysed as described above. For analysis of cleavage of multiple proteins, HEK293T cells were cultured in the presence of azide-modified moieties (myristic acid, palmitic acid, geranylgeranyl alcohol) according to the Click-iT protocol. Cell lysates were incubated with 3 μg of IpaJ Δ50 (wild type or C64A) for 30 min at 37 °C. Proteins containing azide-modified moieties were then labelled with Alex Fluor 647 Alkyne and purified according to the Click-iT labelling protocol.

31. Shima, D. T., Halder, K., Pepperkok, R., Watson, R. & Warren, G. Partitioning of the Golgi apparatus during mitosis in living HeLa cells. *J. Cell Biol.* **137**, 1211–1228 (1997).
32. Steele-Mortimer, O., Meresse, S., Gorvel, J. P., Toh, B. H. & Finlay, B. B. Biogenesis of *Salmonella typhimurium*-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. *Cell. Microbiol.* **1**, 33–49 (1999).
33. Shiloh, M. U. et al. Phenotype of mice and macrophages deficient in both phagocyte oxidase and inducible nitric oxide synthase. *Immunity* **10**, 29–38 (1999).
34. Rivera, V. M. et al. Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science* **287**, 826–830 (2000).
35. Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
36. Alto, N. M. & Dixon, J. E. Analysis of Rho-GTPase mimicry by a family of bacterial type III effector proteins. *Methods Enzymol.* **439**, 131–143 (2008).

# SIRT6 regulates TNF- $\alpha$ secretion through hydrolysis of long-chain fatty acyl lysine

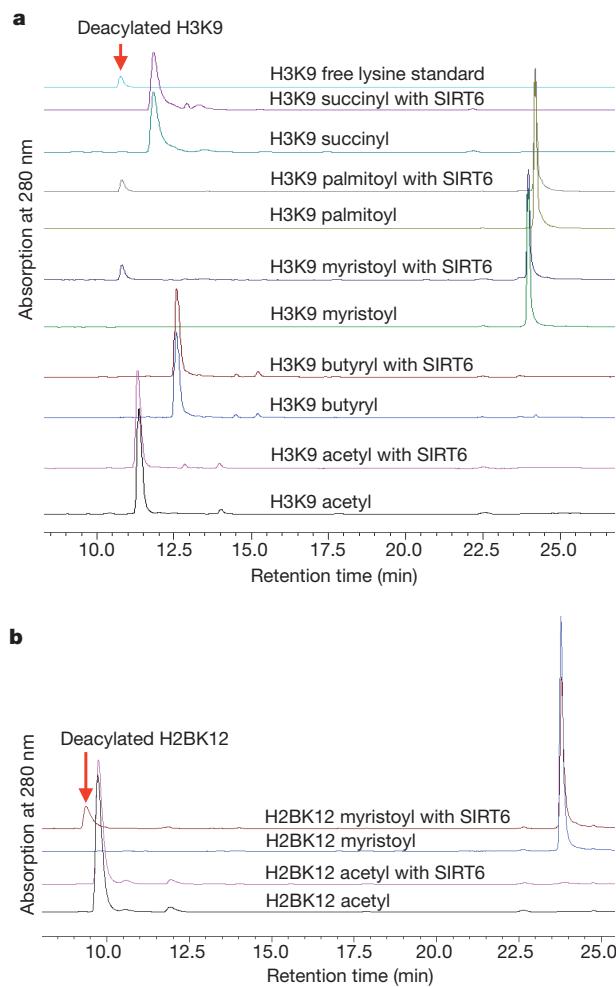
Hong Jiang<sup>1\*</sup>, Saba Khan<sup>1\*</sup>, Yi Wang<sup>2\*</sup>, Guillaume Charron<sup>3</sup>, Bin He<sup>1</sup>, Carlos Sebastian<sup>4</sup>, Jintang Du<sup>1</sup>, Ray Kim<sup>1</sup>, Eva Ge<sup>1</sup>, Raul Mostoslavsky<sup>4</sup>, Howard C. Hang<sup>3</sup>, Quan Hao<sup>2</sup> & Hening Lin<sup>1</sup>

The Sir2 family of enzymes or sirtuins are known as nicotinamide adenine dinucleotide (NAD)-dependent deacetylases<sup>1</sup> and have been implicated in the regulation of transcription, genome stability, metabolism and lifespan<sup>2,3</sup>. However, four of the seven mammalian sirtuins have very weak deacetylase activity *in vitro*. Here we show that human SIRT6 efficiently removes long-chain fatty acyl groups, such as myristoyl, from lysine residues. The crystal structure of SIRT6 reveals a large hydrophobic pocket that can accommodate long-chain fatty acyl groups. We demonstrate further that SIRT6 promotes the secretion of tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) by removing the fatty acyl modification on K19 and K20 of TNF- $\alpha$ . Protein lysine fatty acylation has been known to occur in mammalian cells, but the function and regulatory mechanisms of this modification were unknown. Our data indicate that protein lysine fatty acylation is a novel mechanism that regulates protein secretion. The discovery of SIRT6 as an enzyme that controls protein lysine fatty acylation provides new opportunities to investigate the physiological function of a protein post-translational modification that has been little studied until now.

It was recently demonstrated that Sirt5, one of the four sirtuins with weak deacetylase activity, preferentially hydrolyses succinyl and malonyl lysine<sup>4,5</sup>. The discovery of the novel Sirt5 activity suggested that other sirtuins with weak deacetylase activity may use alternative substrates as well. We therefore set out to investigate whether SIRT6 has any new enzymatic activity. SIRT6 has been reported to be important for DNA repair, transcriptional regulation of genes important for metabolism and immune responses, and for lifespan<sup>6–10</sup>. In most cases, the biological functions have been linked to the sequence-specific deacetylase activity of SIRT6 on histone H3K9 and K56, but not other peptide sequences<sup>9,11,12</sup>. The deacetylase activity of SIRT6 on H3K9 and K56 can indeed be detected in our assay. However, the catalytic efficiency is low (Table 1), indicating that the deacetylase activity may only account for a part of its biological function.

To discover possible novel activity of SIRT6, we synthesized H3K9 peptides with different acyl groups (acetyl, malonyl, succinyl, butyryl, myristoyl and palmitoyl) and assayed these peptides with recombinant SIRT6. SIRT6 can hydrolyse long-chain fatty acyl groups efficiently (Fig. 1a). The  $k_{\text{cat}}/K_m$  for demyristoylation ( $1,400 \text{ s}^{-1} \text{ M}^{-1}$ ) is approximately 300-fold better than that for deacetylation ( $4.8 \text{ s}^{-1} \text{ M}^{-1}$ ). The

increased catalytic efficiency comes mainly from the decrease in  $K_m$ . For deacetylation, the  $K_m$  is  $810 \mu\text{M}$ , but for demyristoylation, the  $K_m$  is  $3.4 \mu\text{M}$ . The activity of SIRT6 (a class IV sirtuin) is similar to that of PfSir2A<sup>5</sup>, which was classified as a Class III sirtuin<sup>13</sup> but lacks the conserved Arg and Tyr residues that recognize negatively charged



**Figure 1 | SIRT6 preferentially hydrolyses long-chain fatty acyl lysine *in vitro*.** **a**, High-performance liquid chromatography traces showing SIRT6-catalysed hydrolysis of different acyl peptides based on the H3K9 sequence. **b**, H2BK12 myristoyl peptide can be hydrolysed by SIRT6, whereas the corresponding acetyl peptide cannot. Reactions were carried out with  $50 \mu\text{M}$  peptide,  $1 \mu\text{M}$  SIRT6,  $20 \text{ mM}$  Tris pH 8.0,  $0.5 \text{ mM}$  NAD and  $1 \text{ mM}$  dithiothreitol (DTT) at  $37^\circ\text{C}$  for 30 min.

**Table 1 | Catalytic efficiencies of SIRT6 on different acyl peptides**

| Acy peptide                 | $k_{\text{cat}} (\text{s}^{-1})$ | $K_m (\mu\text{M})$ | $k_{\text{cat}}/K_m (\text{s}^{-1} \text{ M}^{-1})$ |
|-----------------------------|----------------------------------|---------------------|---|
| H3K9 acetyl                 | $0.0039 \pm 0.0006$              | $810 \pm 160$       | 4.8   |
| H3K9 butyryl                | $0.0021 \pm 0.0004$              | $200 \pm 120$       | 10  |
| H3K9 octanoyl               | $0.0046 \pm 0.0005$              | $40 \pm 10$         | $1.2 \times 10^2$                                   |
| H3K9 myristoyl              | $0.0049 \pm 0.0004$              | $3.4 \pm 0.9$       | $1.4 \times 10^3$                                   |
| H3K9 palmitoyl              | $0.0027 \pm 0.0002$              | $0.9 \pm 0.4$       | $3.0 \times 10^3$                                   |
| TNF- $\alpha$ K19 myristoyl | $0.0020 \pm 0.0002$              | $2.4 \pm 0.6$       | $8.3 \times 10^2$                                   |
| TNF- $\alpha$ K20 myristoyl | $0.0050 \pm 0.0004$              | $4.5 \pm 1.1$       | $1.1 \times 10^3$                                   |

<sup>1</sup>Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. <sup>2</sup>Department of Physiology, University of Hong Kong, Hong Kong, China. <sup>3</sup>The Laboratory of Chemical Biology and Microbial Pathogenesis, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. <sup>4</sup>The Massachusetts General Hospital Cancer Center, Harvard Medical School, 185 Cambridge St., Boston, Massachusetts 02114, USA.

\*These authors contributed equally to this work.

succinyl or malonyl groups<sup>4</sup>. This emphasizes that it is important to examine the activity of each sirtuin experimentally as bioinformatics predictions may not be sufficient.

The conclusion that SIRT6 is better at hydrolysing long-chain fatty acyl groups was also supported by its ability to hydrolyse myristoyl group from other peptide sequences. The deacetylase activity of SIRT6 was reported to be sequence-specific. Only H3K9 and H3K56 acetyl peptides, but not other peptides, were reported to be deacetylated by SIRT6<sup>9,11,12</sup>. To test whether the more efficient demyristoylation activity allows SIRT6 to use more peptide substrates, we synthesized acetyl and myristoyl peptides based on the H2B K12 sequence. Consistent with previous reports<sup>9,11,12</sup>, the hydrolysis of the acetyl peptide by SIRT6 was undetectable. In contrast, the hydrolysis of the corresponding myristoyl peptide could be readily detected (Fig. 1b). These results confirm the preference of SIRT6 for long-chain fatty acyl groups.

To determine the structural basis for enhanced SIRT6 activity with fatty acyl peptides, we obtained a crystal structure of SIRT6 in complex with a H3K9 myristoyl peptide and ADP-ribose (ADPR) at 2.2 Å resolution (PDB 3ZG6, Fig. 2a, Supplementary Fig. 1). The overall structure is similar to the published SIRT6 structure<sup>14</sup> (PDB 3K35) (Supplementary Fig. 1). Residues 2–10 and 166–174 of SIRT6 are visible in the H3K9 myristoyl bound structure, whereas the corresponding regions are missing in the published SIRT6 structure without peptide bound<sup>14</sup> (Supplementary Fig. 2). The way that H3K9 myristoyl and ADPR bind to SIRT6 is similar to that seen in other ternary complex structures of other sirtuins, such as Sir2Tm in complex with the p53 acetyl peptide and NAD (Supplementary Fig. 3)<sup>15</sup>. The peptide interacts with SIRT6 through many hydrogen-bonding interactions. Similar to what was observed with other sirtuins, including Sirt5, most of the hydrogen bonds come from main chain C=O and N-H of the myristoyl peptide, with the only side chain hydrogen-bonding coming from Trp 11 (Supplementary Fig. 4). Therefore, the selectivity for peptide sequences is not high, which is consistent with our enzymology data. The myristoyl group is located in a hydrophobic pocket (Fig. 2b) formed by hydrophobic residues from several flexible loops, including Ala 11, Pro 60, Phe 62, Trp 69, Pro 78, Phe 80, Phe 84, Val 113, Leu 130, Leu 184 and Ile 217. The substrate-binding sites in other available human sirtuin structures do not possess such a big hydrophobic pocket (Supplementary Fig. 5). The structural data are thus consistent with the biochemical data and provide a reasonable explanation for the preference of SIRT6 for long-chain fatty acyl groups.

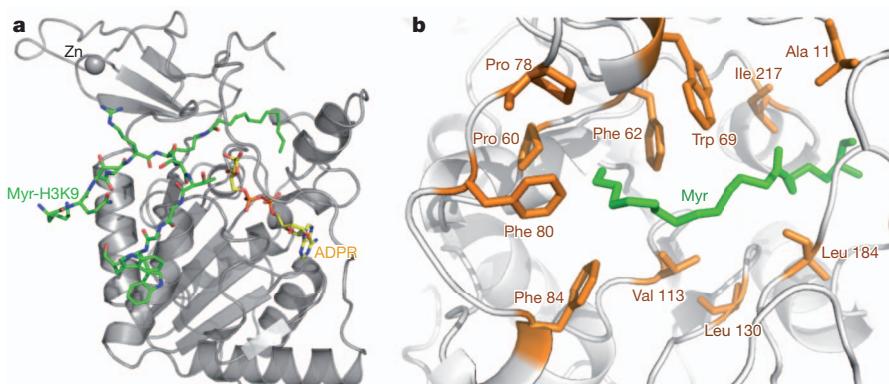
The next important question was whether this activity is physiologically relevant. There are a few proteins known to be modified by fatty acyl groups on lysine, although the function of the modification is unclear<sup>16,17</sup>. One of the fatty acylated proteins is TNF-α<sup>16</sup>, a type II membrane protein with a single transmembrane domain linking the amino-terminal intracellular domain and the carboxy-terminal extracellular domain<sup>18</sup>. When cleaved by proteases, the extracellular domain

is released. The released TNF-α can then bind its receptors and induce various signalling pathways<sup>18</sup>. It was reported that SIRT6 can regulate the synthesis of TNF-α, but the mechanism was unclear<sup>19,20</sup>. We propose that SIRT6 may regulate TNF-α secretion via defatty-acylation.

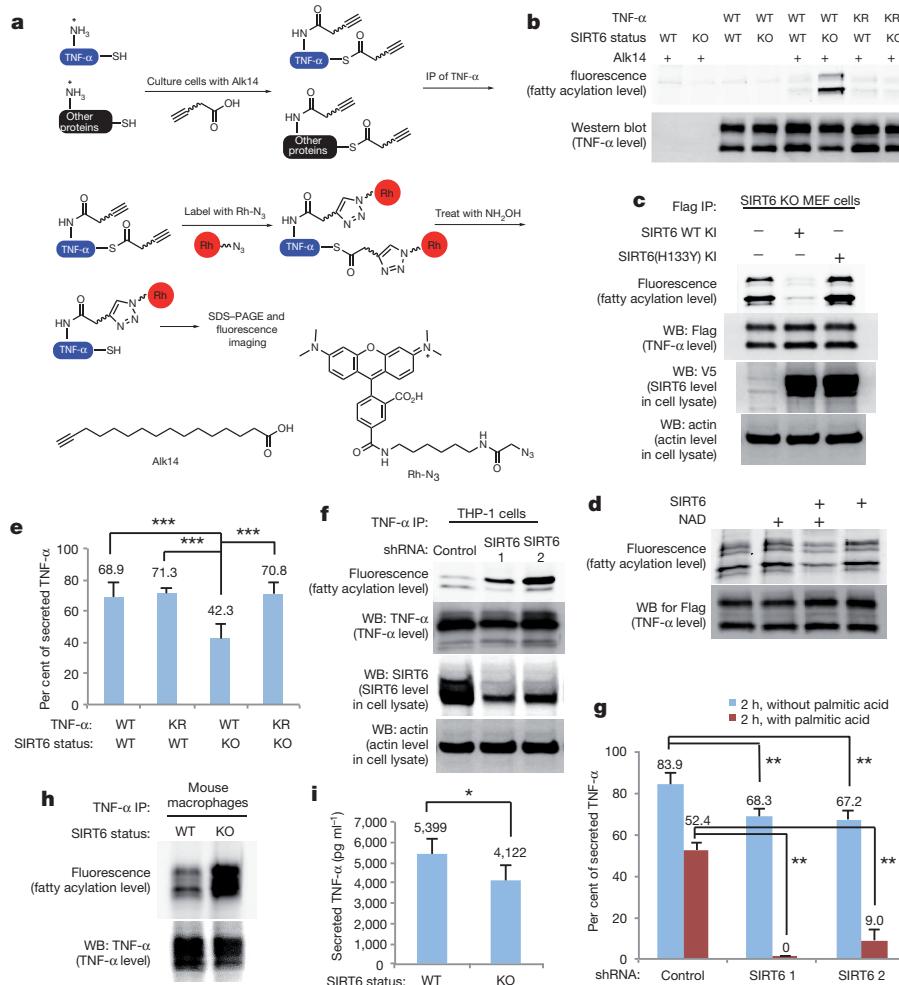
To test this, we measured the fatty acylation level on TNF-α in SIRT6 wild-type (WT) and knockout (KO) mouse embryonic fibroblast (MEF) cells. Flag-tagged TNF-α was transfected into the cells. The cells were cultured in the presence of an alkyne-tagged fatty acid analogue, Alk14, which can covalently label fatty acylated proteins<sup>21,22</sup>. TNF-α was immunoprecipitated and conjugated to rhodamine-azide (Rh-N<sub>3</sub>) using click chemistry (Fig. 3a). A protein will be fluorescently labelled if it is fatty acylated by Alk14. Because Alk14 can also label cysteine residues<sup>22,23</sup>, we treated TNF-α with hydroxylamine to remove cysteine fatty acylation (Fig. 3a). As shown in Fig. 3b, TNF-α from SIRT6 KO MEF cells had significantly increased Alk14 labelling compared to TNF-α from SIRT6 WT MEF cells, indicating that SIRT6 regulates the fatty acylation level on TNF-α. When human SIRT6 WT was overexpressed in SIRT6 KO MEF cells, TNF-α had lower fatty acylation level than TNF-α from cells without overexpression of human SIRT6, whereas overexpression of human SIRT6(H133Y) catalytic mutant did not have much effect on TNF-α fatty acylation (Fig. 3c), indicating that enzymatic activity of SIRT6 is required for controlling TNF-α fatty acylation.

To further confirm that fatty acylation occurs on lysine residues of TNF-α, we mutated K19 and K20, the reported sites of myristoylation on TNF-α<sup>16</sup>, to Arg (labelled as 'KR' mutant in Fig. 3b). The labelling intensities of the mutant from SIRT6 KO MEF cells dropped to background level (Fig. 3b), confirming that K19 and K20 are the major fatty acylation sites. Furthermore, TNF-α isolated from SIRT6 KO MEF cells can be defatty-acylated *in vitro* in an NAD-dependent manner (Fig. 3d). Synthetic TNF-α K19 and K20 myristoyl peptides can be efficiently hydrolysed by SIRT6 (Table 1). To rule out that TNF-α could also be regulated by lysine acetylation, we examined the acetylation level using a pan-specific acetyl lysine antibody. No acetylation was detected on K19 and K20 of TNF-α (Supplementary Fig. 6). Thus, SIRT6 regulates the fatty acylation level, but not acetylation level on K19 and K20 of TNF-α.

We next investigated the function of TNF-α fatty acylation on K19 and K20. It was reported that the amount of TNF-α detected in the media of SIRT6 KO cells is less than that in the media of SIRT6 WT cells<sup>20</sup>. This was attributed to the regulation of TNF-α synthesis by SIRT6<sup>20</sup>. Given that SIRT6 regulates lysine fatty acylation, we propose that SIRT6 may regulate the secretion of TNF-α. To test this, Flag-tagged TNF-α was transiently transfected into SIRT6 WT and KO MEF cells. The amount of secreted TNF-α in the medium and the amount of TNF-α in the cells were measured by ELISA (Supplementary Fig. 7). The percentage of secreted TNF-α was then calculated. TNF-α secretion efficiency was lower in SIRT6 KO MEF cells than in



**Figure 2 | Structure basis for SIRT6 activity with long-chain fatty acyl groups.** **a**, Overall structure of SIRT6 with myristoyl H3K9 (Myr-H3K9, green) peptide and ADP-ribose (ADPR, yellow) bound. **b**, Hydrophobic residues in SIRT6 that accommodate the myristoyl (Myr) group.



**Figure 3 | SIRT6 regulates TNF- $\alpha$  fatty acylation and secretion. a**, Method of using Alk14 to detect TNF- $\alpha$  fatty acylation. **b**, SIRT6 controls TNF- $\alpha$  fatty acylation on K19 and K20. KR, TNF- $\alpha$ (K19R, K20R) mutant. **c**, H133 of SIRT6 is required for TNF- $\alpha$  defatty-acylation. KI, knock-in. **d**, SIRT6 defatty-acylates TNF- $\alpha$  *in vitro*. **e**, SIRT6 regulates secretion of TNF- $\alpha$  in MEF cells. *n* = 6;

SIRT6 WT MEF cells (Fig. 3E). The secretion efficiency is due to lysine fatty acylation because the secretion efficiency of TNF- $\alpha$  KR mutant was not affected by SIRT6 knockout. These data indicate that lysine fatty acylation regulates TNF- $\alpha$  secretion and SIRT6 promotes TNF- $\alpha$  secretion by defatty-acylation.

We further investigated whether endogenous TNF- $\alpha$  are also regulated by SIRT6. For this purpose, we used both human THP-1 cells and bone-marrow-derived mouse macrophages. Two SIRT6 knockdown (KD) THP-1 cell lines and one control KD THP-1 cell line were generated (Fig. 3f). TNF- $\alpha$  from SIRT6 KD cells contained more fatty acylation than TNF- $\alpha$  from the control KD cells (Fig. 3f). The amount of secreted TNF- $\alpha$  in the medium and the amount of TNF- $\alpha$  in the cells were measured (Supplementary Fig. 8) and the percentage of secreted TNF- $\alpha$  was calculated (Fig. 3g). The secretion efficiency of TNF- $\alpha$  was lower in SIRT6 KD THP-1 cells than in control KD cells, especially when the cells were supplemented with palmitic acid (Fig. 3g). Similar results were obtained for TNF- $\alpha$  in mouse macrophages. TNF- $\alpha$  in SIRT6 WT macrophages had lower lysine fatty acylation than in SIRT6 KO macrophages (Fig. 3h), whereas TNF- $\alpha$  secretion efficiency in SIRT6 WT macrophages was higher than in SIRT6 KO macrophages (Fig. 3i).

SIRT6 was reported to be mainly localized in the nucleus<sup>9</sup>. The regulation of TNF- $\alpha$  secretion by SIRT6 indicated that SIRT6 might be present in secretory organelles, such as the endoplasmic reticulum. We indeed detected SIRT6 in the endoplasmic reticulum fraction of

\*\*\*P < 0.001. **f-i**, SIRT6 regulates fatty acylation level and secretion of endogenous TNF- $\alpha$  in THP-1 cells (**f** and **g**; *n* = 3; \*\*P < 0.01) and bone-marrow-derived macrophages (**h** and **i**; *n* = 5; \*P < 0.02). Secretion data were expressed as mean  $\pm$  s.d. IP, immunoprecipitation; shRNA, short hairpin RNA; WB, western blot.

MEF and THP-1 cell lysates (Supplementary Fig. 9). In SIRT6 KD THP-1 cells or SIRT6 KO MEF cells, less or no SIRT6 was detected in the endoplasmic reticulum fraction compared with SIRT6 WT cells (Supplementary Fig. 9). The presence of SIRT6 in the endoplasmic reticulum provided further support for the conclusion that SIRT6 regulates TNF- $\alpha$  secretion through defatty-acylation of TNF- $\alpha$ .

In summary, our enzymological and structural studies show that human SIRT6, which has weak deacetylase activities *in vitro*, catalyses the hydrolysis of fatty acyl lysine modifications more efficiently than deacetylation. SIRT6 regulates the fatty acylation level on K19 and K20 of TNF- $\alpha$  and this modulates the secretion of TNF- $\alpha$ . It has been reported that SIRT6 regulates the acetylation level on histone H3K9 and K56<sup>9,11,12</sup>. Our results are not in conflict with these reports because it is possible that the association with chromatin can regulate the activity of SIRT6. Alternatively, fatty acylation of lysine residues may occur to histones<sup>22,23</sup>, and SIRT6 can regulate both acetylation and fatty acylation of histones. The significance of our finding is several fold. First, it reveals a novel physiological activity for SIRT6. Second, it demonstrates, as far as we know for the first time, that lysine fatty acylation, a protein post-translational modification that has been little studied until now, has an important role in regulating protein secretion. Other proteins (Supplementary Fig. 10), such as insulin-like growth factor 1 (IGF1) may be regulated by similar mechanisms<sup>6,24</sup>. Third, the discovery of SIRT6 activity on protein lysine fatty acylation provides an avenue to further investigate this under-recognized

protein modification<sup>16,17</sup> and may reveal interesting connections to the well-known cysteine palmitoylation and N-terminal glycine myristylation<sup>25</sup>.

## METHODS SUMMARY

**Expression, purification and crystallization.** Human SIRT6 was overexpressed in *Escherichia coli* and purified using Ni-NTA affinity chromatography, ion exchange and gel filtration. Crystals were grown from 12% PEG6K, 0.1 M MES (2-(N-morpholino)ethanesulphonic acid), pH 6.5.

**Data collection and structure determination.** Data were collected at the Shanghai Synchrotron Radiation Facility (SSRF). The structure was determined by molecular replacement.

**Activity assay of SIRT6 on different acyl peptides.** Acyl peptides were synthesized as previously described<sup>5</sup>. The hydrolysis of different acyl peptides catalysed by SIRT6 was monitored by high-performance liquid chromatography. Kinetic parameters were determined by varying the concentrations of the acyl peptides.

**Detection of long-chain fatty acylation on TNF- $\alpha$ .** Flag-tagged TNF- $\alpha$  was transiently transfected into either SIRT6 WT or KO MEF cells<sup>7</sup>. Endogenous TNF- $\alpha$  in THP-1 cells and SIRT6 WT or KO mouse macrophages was stimulated by LPS. The fatty acyl lysine modification was detected using alkyne-tagged fatty acid analogue Alk14 using a procedure modified from a previously published method<sup>21</sup>.

**Full Methods** and any associated references are available in the online version of the paper.

Received 17 July 2012; accepted 22 February 2013.

1. Imai, S.-i., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
2. Sauve, A. A., Wolberger, C., Schramm, V. L. & Boeke, J. D. The biochemistry of sirtuins. *Annu. Rev. Biochem.* **75**, 435–465 (2006).
3. Michan, S. & Sinclair, D. Sirtuins in mammals: insights into their biological function. *Biochem. J.* **404**, 1–13 (2007).
4. Du, J. et al. Sirt5 is an NAD-dependent protein lysine demalonylase and desuccinylase. *Science* **334**, 806–809 (2011).
5. Zhu, A. Y. et al. *Plasmodium falciparum* Sir2A preferentially hydrolyzes medium and long chain fatty acyl lysine. *ACS Chem. Biol.* **7**, 155–159 (2012).
6. Mostoslavsky, R. et al. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).
7. Zhong, L. et al. The histone deacetylase Sirt6 regulates glucose homeostasis via Hif1 $\alpha$ . *Cell* **140**, 280–293 (2010).
8. Xiao, C. et al. SIRT6 deficiency results in severe hypoglycemia by enhancing both basal and insulin-stimulated glucose uptake in mice. *J. Biol. Chem.* **285**, 36776–36784 (2010).
9. Michishita, E. et al. SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature* **452**, 492–496 (2008).
10. Kanfi, Y. et al. The sirtuin SIRT6 regulates lifespan in male mice. *Nature* **483**, 218–221 (2012).
11. Yang, B., Zwaans, B. M. M., Eckersdorff, M. & Lombard, D. B. The sirtuin SIRT6 deacetylates H3 K56Ac *in vivo* to promote genomic stability. *Cell Cycle* **8**, 2662–2663 (2009).
12. Michishita, E. et al. Cell cycle-dependent deacetylation of telomeric histone H3 lysine K56 by human SIRT6. *Cell Cycle* **8**, 2664–2666 (2009).
13. Frye, R. A. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochem. Biophys. Res. Commun.* **273**, 793–798 (2000).
14. Sebastián, C. et al. Deacetylase activity is required for STAT5-dependent GM-CSF functional activity in macrophages and differentiation to dendritic cells. *J. Immunol.* **180**, 5898–5906 (2008).
15. Hoff, K. G., Avalos, J. L., Sens, K. & Wolberger, C. Insights into the sirtuin mechanism from ternary complexes containing NAD<sup>+</sup> and acetylated peptide. *Structure* **14**, 1231–1240 (2006).
16. Stevenson, F. T. et al. The 31-kDa precursor of interleukin 1 alpha is myristoylated on specific lysines within the 16-kDa N-terminal propiece. *Proc. Natl. Acad. Sci. USA* **90**, 7245–7249 (1993).
17. Stevenson, F. T., Bursten, S. L., Locksley, R. M. & Lovett, D. H. Myristyl acylation of the tumor necrosis factor alpha precursor on specific lysine residues. *J. Exp. Med.* **176**, 1053–1062 (1992).
18. Locksley, R. M., Killeen, N. & Lenardo, M. J. The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**, 487–501 (2001).
19. Bruzzone, S. et al. Catastrophic NAD depletion in activated T lymphocytes through Nampt inhibition reduces demyelination and disability in EAE. *PLoS ONE* **4**, e7897 (2009).
20. Van Gool, F. et al. Intracellular NAD levels regulate tumor necrosis factor protein synthesis in a sirtuin-dependent manner. *Nature Med.* **15**, 206–210 (2009).
21. Charron, G. et al. Robust fluorescent detection of protein fatty-acylation with chemical reporters. *J. Am. Chem. Soc.* **131**, 4967–4975 (2009).
22. Wilson, J. P. et al. Proteomic analysis of fatty-acylated proteins in mammalian cells with chemical reporters reveals S-acylation of histone H3 variants. *Mol. Cell. Proteomics* **10**, M110.001198 (2011).
23. Martin, B. R. & Cravatt, B. F. Large-scale profiling of protein palmitoylation in mammalian cells. *Nature Methods* **6**, 135–138 (2009).
24. Schwer, B. et al. Neural sirtuin 6 (Sirt6) ablation attenuates somatic growth and causes obesity. *Proc. Natl. Acad. Sci. USA* **107**, 21790–21794 (2010).
25. Walsh, C. T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory* (Roberts, 2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported in part by NIH R01GM086703 (H.L.), R01GM093072 (R.M.), Hong Kong GRF766510 (Q.H.) and NIH R01GM087544 (H.C.H.). We thank C. Zhang for help with the cloning of SIRT6 WT and H133Y to generate lentiviral particles and the staff at the Shanghai Synchrotron Radiation Facility for assistance during the data collection.

**Author Contributions** H.J. designed and carried out all biochemical experiments involving TNF- $\alpha$  and synthesized Rh-N<sub>3</sub>. S.K. synthesized acyl peptides and carried out all enzymology experiments of SIRT6. Y.W. carried out all crystallography experiments. G.C. and B.H. synthesized Alk14. G.C. and H.C.H. provided expertise on the labelling experiments using Alk14. C.S. and R.M. generated the MEF cells and bone marrow-derived macrophages from SIRT6 WT and KO mice. J.D., R.K. and E.G. contributed to the cloning, expression and purification of SIRT6. Q.H. directed the structural studies and wrote the structural part of the manuscript. H.L. directed the biochemical studies, coordinated the collaborations among different labs, and wrote the manuscript with help from H.J., S.K., Y.W., R.M., H.C.H. and Q.H.

**Author Information** The crystal structure of SIRT6 in complex with a H3K9 myristoyl peptide and ADP-ribose is deposited in the Protein Data Bank as accession number 3ZG6. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.L. (hl379@cornell.edu) and Q.H. (qhao@hku.hk).

## METHODS

**Reagents.** Mouse monoclonal anti-Flag M2 antibody conjugated with horseradish peroxidase, anti-Flag M2 affinity gel, and human/mouse SIRT6 antibody (S4322) were from Sigma. Human TNF- $\alpha$  antibody (D5G9) and mouse TNF- $\alpha$  antibody (D2D4) were from Cell Signaling Technology. TNF- $\alpha$  Affibody immobilized on agarose (ab31909) and human SIRT6 antibody (ab88494) were from Abcam. Lamin A/C antibody (636),  $\beta$ -actin antibody (C4) and GRP 78 antibody (H-129) were from Santa Cruz Biotechnology. The rabbit pan-specific anti-acetyl lysine antibody was from ImmuneChem Pharmaceuticals. The goat anti-rabbit and mouse IgG conjugated with horseradish peroxidase and protein A/G plus-agarose were from Santa Cruz Biotechnology. Human and mouse TNF- $\alpha$  ELISA kits were from eBioscience. Brefeldin A (BFA), phorbol 12-myristate 13-acetate (PMA), lipopolysaccharides from *Escherichia coli* 0111:B4 (LPS) and palmitic acid were purchased from Sigma. Alk14 was synthesized according to reported procedures<sup>21</sup>. SIRT6 WT or KO MEF cells were generated as previously reported<sup>7</sup>. THP-1 cells were purchased from ATCC.

### Cloning, expression and purification of full-length SIRT6 for activity assay.

The open reading frame of full-length human SIRT6 (1–355) was inserted into a pET28a vector between the BamHI and NotI sites. This plasmid was transformed into *E. coli* Arctic Express (DE3) cells. The cells were cultured at 37 °C in 2 × YT culture medium (5 g of NaCl, 16 g of bactotryptone and 10 g of yeast extract per litre). Isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG, 0.2 mM) was used to induce expression when attenuation  $D_{600\text{nm}}$  was 0.6, and the culture was grown for 20 h at 289 K. Cells were collected by centrifugation at 7,330g for 10 min and then re-suspended in lysis buffer (20 mM Tris-HCl, pH 7.2, 500 mM NaCl and 2% glycerol). Cells were lysed using a cell disrupter. After centrifugation at 29,300g for 25 min at 277 K, the supernatant was loaded onto a nickel column (QFF-Sepharose, Amersham Biosciences) pre-equilibrated with 20 mM Tris-HCl pH 7.2 with 500 mM NaCl. The protein was eluted with a linear gradient of imidazole (0–500 mM). The desired fractions were pooled, concentrated and buffer exchanged to cation exchange buffer (20 mM Tris pH 7.2, 80 mM NaCl, 5% glycerol). The protein was loaded onto a cation exchange column (Amersham Biosciences) and eluted with 1 M NaCl, 20 mM Tris-HCl, pH 7.2, 2% glycerol. The purified protein was stored at –80 °C.

### Cloning, expression and purification of truncated SIRT6 for crystallization.

Human SIRT6 (1–294) was inserted into a pET28a vector between the NdeI and NotI sites. The protein was overproduced at 37 °C in *E. coli* Rosetta (DE3) strain using 2 × YT medium. The expression and purification method was the same as that used for the full-length SIRT6 except that after Ni column purification the protein was further purified by gel filtration using a Superdex75 column (Amersham Biosciences). The protein was eluted with 20 mM Tris-HCl, pH 7.0, 100 mM NaCl. After concentration to 8 mg ml<sup>–1</sup>, the target protein was frozen at –80 °C.

**Synthesis of acyl peptides.** The peptides were synthesized and purified as described previously<sup>5</sup>. The identity of the peptides was confirmed using LCQ Fleet ThermoFisher Mass Spectrometer. The acetyl, butyryl peptides were dissolved in 25% (v/v) DMSO in water. The longer chain fatty acyl peptides were dissolved in DMSO. The concentrations of peptides were determined at 280 nm using extinctions coefficient of the two tryptophans attached at the C termini of the peptides.

**Deacylation activity assay.** The activity of SIRT6 was analysed using reverse phase high-performance liquid chromatography on a Kinetex XB-C18 column (100A, 75 mm × 4.60 mm, 2.6  $\mu$ m, Phenomenex). SIRT6 full length (1  $\mu$ M) was incubated in a reaction mixture (60  $\mu$ l) containing 20 mM Tris pH 8.0, 1 mM DTT, 0.5 mM NAD and 50  $\mu$ M H3K9 acyl peptides at 37 °C for 30 min. Total DMSO content in the reaction was <2.5% unless mentioned otherwise. The reactions were quenched with 60  $\mu$ l of 0.5 N HCl in methanol. The reactions were then monitored by high-performance liquid chromatography as described later in the kinetics assay.

**Kinetics assay for acetyl and butyryl peptides.** Peptide concentrations were varied from 0 to 250  $\mu$ M for H3K9 butyryl peptide and from 0 to 600  $\mu$ M for the acetyl peptide. The reaction mixtures (60  $\mu$ l with 2 mM NAD, 1 mM DTT, 20 mM Tris, pH 8.0, 4  $\mu$ M SIRT6 full length, and acyl peptides at various concentrations) were incubated for 30 min at 37 °C. The reaction was stopped using 60  $\mu$ l of 0.5 N HCl in methanol. The reaction mixtures were spun at 18,000g for 10 min and were analysed on a Kinetex XB-C18 column (100A, 100 mm × 4.60 mm, 2.6  $\mu$ m, Phenomenex). The gradient of 20–40% B (acetonitrile with 0.1% TFA) in 17 min at 0.5 ml min<sup>–1</sup> was used.

**Kinetics assay for long-chain fatty acyl peptides.** The peptide concentrations were varied from 1 to 20  $\mu$ M. The reactions (60  $\mu$ l with 2 mM NAD, 1 mM DTT, 20 mM Tris pH 8.0, 0.2  $\mu$ M recombinant SIRT6 full length, and acyl peptides at different concentrations) was incubated for 15 min at 37 °C. The reaction was stopped using 60  $\mu$ l of 0.5N HCl in methanol. The reaction mixtures were spun

at 18,000g for 10 min and were analysed on a Kinetex XB-C18 column (100A, 75 mm × 4.60 mm, 2.6  $\mu$ m, Phenomenex). The gradient of 0–55% B in 10 min at 0.5 ml min<sup>–1</sup> was used. The product and the substrate peaks were quantified using absorbance at 280 nm and converted to initial rates, which were then plotted against the acyl peptide concentrations and fitted using Kaleidagraph.

**Crystallization, X-ray data collection and structure determination.** Crystals of complex SIRT6 with H3K9-Myr and ADPR were obtained by hanging drop vapour-diffusion method at 291 K using commercial screens from Hampton Research. Each drop consisting of 1  $\mu$ l of 10 mg ml<sup>–1</sup> protein complex solution (20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 5 mM DTT) and 1  $\mu$ l reservoir solution was equilibrated against 400  $\mu$ l reservoir solution. The qualified crystals of SIRT6 grew with a cube profile within 1 week with a reservoir containing 12% PEG6K, 0.1 M MES, pH 6.5. The mixture of 30% glycerol with reservoir solution above was used as cryogenic liquor. The X-ray diffraction data were collected at 100 K in a liquid nitrogen gas stream at the Shanghai Synchrotron Radiation Facility BL17U. 180 frames were collected with a 1° oscillation and the data were indexed and integrated using the program HKL2000<sup>26</sup>. The complex structure of SIRT6 with H3K9 myristoyl peptide and ADPR was solved by molecular replacement using the program Molrep from the CCP4 Suite<sup>27</sup> with the published SIRT6 structure (PDB: 3K35)<sup>28</sup> as the search model. Refinement and model building were performed with REFMAC5 and COOT from CCP4. The X-ray diffraction data collection and structure refinement statistics are shown in Supplementary Table 1.

**Generation of SIRT6 KO MEF cells with human SIRT6 WT and H133Y mutant knock-in.** Human SIRT6 WT or H133Y complementary DNA was inserted to lentiviral vector (engineered from pSIN-EF2-OCT4 vector with OCT4 deletion to convert into gateway destination vector, and provided by C. Zhang) using Gateway Cloning. After co-transfection of SIRT6 WT or H133Y lentiviral plasmid, pCMV-dR8.2, and pMD2.G into 293T cells, the medium was collected to infect SIRT6 KO MEF cells. The SIRT6 WT or H133Y knock-in cells were selected using 1.5  $\mu$ g ml<sup>–1</sup> puromycin in complete cell culture medium. Overexpressed SIRT6 WT and H133Y mutant have a C-terminal V5 tag.

**Cloning of TNF- $\alpha$ .** For construction of human TNF- $\alpha$  expression vector to express wild-type TNF- $\alpha$  protein (TNF- $\alpha$  WT) with N-terminal Flag tag and C-terminal haemagglutinin tag, human full-length TNF- $\alpha$  cDNA were generated by PCR and inserted into pCMV4A vector at EcoRV and XhoI sites. The plasmid of TNF- $\alpha$  double lysine mutant (K19R, K20R), TNF- $\alpha$  KR, was made by overlap extension PCR.

**Transfection of TNF- $\alpha$  into MEF cells.** SIRT6 WT and KO MEF cells were maintained in DMEM medium containing glucose and L-glutamine (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (Invitrogen). The pCMV vectors containing target genes were transfected into cells using FuGene 6 (Promega) according to the manufacturer's protocol. Empty pCMV vector was transfected into cells as negative control. This transfection method was also applied to SIRT6 KO MEF cells with human SIRT6 WT or H133Y mutant knock-in for TNF- $\alpha$  overexpression.

**Labelling of TNF- $\alpha$  in MEF cells with Alk14.** SIRT6 WT or KO MEF cells were cultured with fresh medium containing 20  $\mu$ M Alk14 and 5  $\mu$ g ml<sup>–1</sup> brefeldin A for 12 h after transient transfection of TNF- $\alpha$ . Cells without Alk14 or without overexpression of TNF- $\alpha$  were used as negative controls. Cells were collected by cell scraper and resuspended in 1 × PBS buffer. After centrifugation at 500g for 5 min at 4 °C, the cell pellet was dissolved in lysis buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 10% glycerol, 1% Nonidet P-40) with protease inhibitor cocktail (Sigma). The supernatant was collected after centrifugation at 16,000g for 10 min at 4 °C, and used for western blotting, immunoprecipitation or TNF- $\alpha$  ELISA experiment.

**Generation of stable SIRT6 KD THP-1 cells.** SIRT6 shRNA lentiviral plasmids in pLKO.1-puro vector were purchased from Sigma. SIRT6 shRNA 1 (TRCN0000378253) cccggcgtacgcgtccggagacacgtctcgaggactgtgcgtggacgtactgttttgg and shRNA 2 (TRCN0000232528) cccggaaatagtgccaatgtaaatgcgttgacacttggcatttcctttttt were used. After co-transfection of SIRT6 shRNA plasmid, pCMV-dR8.2 and pMD2.G into 293T cells, the medium was collected to infect THP-1 cells. The SIRT6 KD cells were selected using 1.5  $\mu$ g ml<sup>–1</sup> puromycin in complete cell culture medium. THP-1 cells infected with lentivirus containing control shRNA plasmid were carried out similarly.

**Labelling of TNF- $\alpha$  in THP-1 cells with Alk14.** THP-1 cells (with control shRNA, SIRT6 shRNA 1 or SIRT6 shRNA 2, at 0.6 × 10<sup>6</sup> cells ml<sup>–1</sup>) were treated with PMA (200 ng ml<sup>–1</sup>) for 24 h. Then cells were cultured with fresh medium containing Alk14 (50  $\mu$ M) and LPS (1  $\mu$ g ml<sup>–1</sup>). After 1 h, brefeldin A was added to culture medium (5  $\mu$ g ml<sup>–1</sup>). Cells were collected after 10 h, and dissolved in lysis buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 10% glycerol, 1% Nonidet P-40) with protease inhibitor cocktail (Sigma). The supernatant was collected after centrifugation at 16,000g for 10 min at 4 °C and used for western blotting or immunoprecipitation.

**Isolation of SIRT6 WT and KO macrophages and TNF- $\alpha$  secretion.** Bone-marrow-derived macrophages were isolated from SIRT6 WT and KO mice as described<sup>14</sup> with some modifications. Briefly, femurs and tibiae were dissected and marrow tissue eluted by irrigation with DMEM. Cells were cultured in Petri dishes containing complete media (DMEM + 20% FCS + 10 ng ml<sup>-1</sup> recombinant murine M-CSF + penicillin/streptomycin) for 7–10 days at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere. For TNF- $\alpha$  mRNA expression analysis, 1.5 × 10<sup>5</sup> macrophages were plated in six-well plates containing complete media and incubated overnight. Next day, LPS (100 ng ml<sup>-1</sup>) was added and the cells were cultured for 12 h. For TNF- $\alpha$  protein expression and secretion, cells were plated as above and, after 12 h of LPS stimulation, the supernatants were collected and used for ELISA analysis. To measure TNF- $\alpha$  fatty acylation levels, 1.5 × 10<sup>6</sup> macrophages were plated in 6-cm plates in complete media and incubated overnight. Next day, cells were treated using the same procedure as described above for THP-1 cells.

**Western blot for TNF- $\alpha$ .** Protein samples were separated by 18% SDS-PAGE and transferred to PVDF membrane. The membrane was blocked with 5% BSA in TBST (25 mM Tris, pH 7.4, 150 mM NaCl, 0.1% Tween-20), incubated with antibodies in TBST, and developed in ECL Plus western blotting detection reagents (GE Healthcare). The chemiluminescence signal was recorded by Storm 860 Imager (Amersham Biosciences) and analysed with ImageQuant TL v2005.

**Immunoprecipitation of Flag-TNF- $\alpha$  using anti-Flag affinity gel.** Cell lysate (250–400 µg from SIRT6 WT or KO MEF cells) was incubated with 20 µl suspension of anti-Flag M2 affinity gel for 2 h at 4 °C. After centrifugation at 500g for 2 min at 4 °C, the affinity gel was washed three times with 500 µl washing buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 0.2% Nonidet P-40) and used for later experiments.

**Immunoprecipitation of endogenous TNF- $\alpha$  from THP-1 cells.** Cell lysate (~500 µg) was incubated with 20 µl suspension of TNF- $\alpha$  Affibody immobilized agarose for 4 h at 4 °C. After centrifugation at 500g for 2 min at 4 °C, the agarose was washed three times with 500 µl washing buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 0.2% Nonidet P-40) and used for later experiments.

**Immunoprecipitation of endogenous TNF- $\alpha$  from bone-marrow-derived mouse macrophage.** Cell lysate (~100 µg from SIRT6 WT or KO macrophages) was incubated with 2 µg mouse TNF- $\alpha$  antibody for 2 h at 4 °C. Then the mixture was incubated with 40 µl suspension of protein A/G plus-agarose for 8 h at 4 °C. After centrifugation at 500g for 2 min at 4 °C, the agarose was washed three times with 500 µl washing buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 0.2% Nonidet P-40) and used for later experiments.

**Detection of fatty acylation on TNF- $\alpha$  by fluorescence.** After immunoprecipitation, the gel/agarose was re-suspended in 10 µl buffer (25 mM Tris, pH 7.4, 50 mM NaCl, 1% Nonidet P-40) for click chemistry. Rh-N<sub>3</sub> (in DMF) was added to the above suspension to a final concentration of 200 µM, followed by the addition of Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (in DMF, final concentration 600 µM), CuSO<sub>4</sub> (in water, final concentration 2 mM), and TCEP (in water, final concentration 2 mM). The click chemistry was allowed to proceed at room temperature for 60 min. The reaction mixture was mixed with protein loading buffer (final 2 ×) and heated at 95 °C for 10 min. After centrifugation at 16,000g for 2 min at room temperature, the supernatant was collected and heated with hydroxylamine (pH 7.2, 60 mM) at 95 °C for 7 min. The samples were resolved by SDS-PAGE using 18% acrylamide gel. Rhodamine fluorescence signal was recorded by Typhoon 9400 Variable Mode Imager (GE Healthcare Life Sciences) with setting of Green (532 nm)/580BP30 PMT 500 V (normal sensitivity), and analysed by ImageQuant TL v2005.

**Detection of acetyl-lysine on TNF- $\alpha$  by western blot using anti-acetyl lysine antibody.** SIRT6 KO MEF cell lysate (~30 mg), with/without overexpression of

TNF- $\alpha$  WT or TNF- $\alpha$  KR, was incubated with 40 µl suspension of anti-Flag M2 affinity gel for 2 h at 4 °C. After centrifugation at 500g for 2 min at 4 °C, the affinity gel was washed three times with 500 µl washing buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 0.2% Nonidet P-40) and then heated with 2 × protein loading buffer at 100 °C for 10 min. The samples were resolved by SDS-PAGE using 18% acrylamide gel and examined by western-blotting using anti-acetyl-lysine antibody. Acetylated BSA was used for positive control to demonstrate acetyl-lysine signal by western-blotting. After recording acetyl-lysine signal, PVDF membrane was washed with water and stained with Coomassie blue to detect TNF- $\alpha$  protein. Another western-blotting using anti-Flag antibody was also carried out to demonstrate the equal loading of TNF- $\alpha$  WT and TNF- $\alpha$  KR.

**Defatty-acylation of TNF- $\alpha$  by SIRT6 *in vitro*.** TNF- $\alpha$  WT was immunoprecipitated from lysate of SIRT6 KO MEF cells (~1 mg) overexpressing Flag-TNF- $\alpha$  and cultured with Alk14 using 40 µl suspension of anti-Flag M2 affinity gel following procedure describe above. After washing, the gel was divided into four equal aliquots. Each aliquot was re-suspended in 10 µl assay buffer (50 mM Tris, pH 7.4, 100 mM NaCl, 2 mM MgCl<sub>2</sub>, 1% Nonidet P-40) with 20 µM of SIRT6 or BSA and with or without 2 mM NAD and incubated at 37 °C for 1 h. Then detection of fatty acylation by fluorescence after hydroxylamine treatment was carried out with each aliquot as described earlier.

**Secretion of TNF- $\alpha$  and TNF- $\alpha$  KR from SIRT6 WT and KO MEF Cells.** After transient transfection of TNF- $\alpha$  WT or TNF- $\alpha$  KR, SIRT6 WT and KO MEF cells were further incubated with fresh medium for 12 h. Then culture medium and cells were collected separately for detection of TNF- $\alpha$  by using human TNF- $\alpha$  ELISA kit. Secretion efficiency was calculated by TNF- $\alpha$  amount in culture medium versus total amount of TNF- $\alpha$  in culture medium and cells. Six independent experiments were carried out.

**TNF- $\alpha$  secretion from THP-1 cells.** Stable shRNA-infected THP-1 cells (at 0.6 × 10<sup>6</sup> cells ml<sup>-1</sup>) were treated with PMA (200 ng ml<sup>-1</sup>) for 20 h. Treatment without palmitic acid: cells were incubated with fresh medium containing LPS (100 ng ml<sup>-1</sup>). After 2 h, culture medium and cells were collected for human TNF- $\alpha$  ELISA assay. Treatment with palmitic acid: cells were incubated with palmitic acid (final 50 µM) in culture medium for 2 h, and then with fresh medium containing LPS (100 ng ml<sup>-1</sup>) and palmitic acid (50 µM). After 2 h, culture medium and cells were collected for human TNF- $\alpha$  ELISA assay. Three independent experiments were carried out.

**Endoplasmic reticulum fractions from THP-1 cells and SIRT6 WT/KO MEF cells.** Endoplasmic reticulum fraction was obtained from THP-1 cells, SIRT6 WT/KO MEF cells, or SIRT6 KO MEF cells with human SIRT6 WT or H133Y knock-in, using an isopycnic flotation method<sup>29</sup>. Endoplasmic reticulum fraction was solubilized in lysis buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 10% glycerol, 1% Nonidet P-40) with protease inhibitor cocktail (Sigma). The supernatant was collected after centrifugation at 16,000g for 10 min at 4 °C, and used for western blotting.

**Statistical analysis.** Data were expressed as mean ± s.d. (standard deviation, shown as error bars). Differences were examined by two-tailed Student's *t*-test between two groups; and \**P* < 0.02, \*\**P* < 0.01, \*\*\**P* < 0.001.

26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
27. Collaborative Computational Project, number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Pan, P. W. *et al.* Structure and biochemical functions of SIRT6. *J. Biol. Chem.* **286**, 14575–14587 (2011).
29. Stephens, S. B. *et al.* Analysis of mRNA partitioning between the cytosol and endoplasmic reticulum compartments of mammalian cells. *Methods Mol. Biol.* **419**, 197–214 (2008).

# Mechanistic studies of an unprecedented enzyme-catalysed 1,2-phosphono-migration reaction

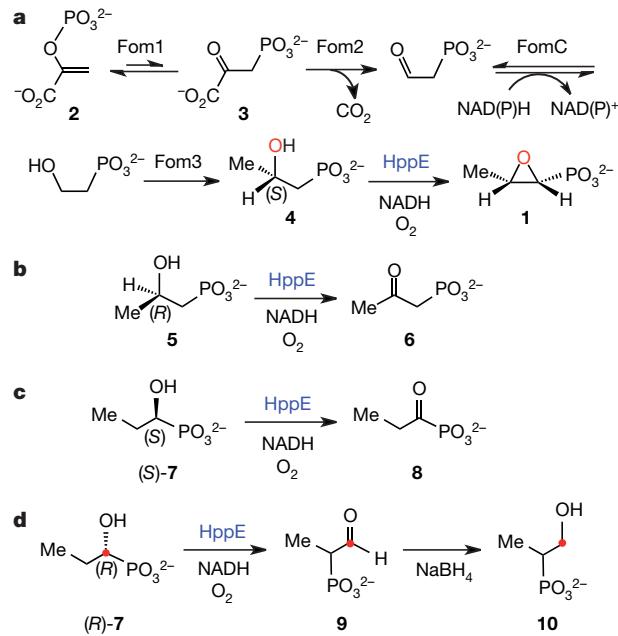
Wei-chen Chang<sup>1†</sup>, Mishti Dey<sup>2\*†</sup>, Pinghua Liu<sup>1\*†</sup>, Steven O. Mansoorabadi<sup>1\*†</sup>, Sung-Ju Moon<sup>1†</sup>, Zongbao K. Zhao<sup>1†</sup>, Catherine L. Drennan<sup>2</sup> & Hung-wen Liu<sup>1</sup>

**(S)-2-hydroxypropylphosphonate ((S)-2-HPP) epoxidase (HppE)** is a mononuclear non-haem-iron-dependent enzyme<sup>1–3</sup> responsible for the final step in the biosynthesis of the clinically useful antibiotic fosfomycin<sup>4</sup>. Enzymes of this class typically catalyse oxygenation reactions that proceed via the formation of substrate radical intermediates. By contrast, HppE catalyses an unusual dehydrogenation reaction while converting the secondary alcohol of (S)-2-HPP to the epoxide ring of fosfomycin<sup>1,5</sup>. Here we show that HppE also catalyses a biologically unprecedented 1,2-phosphono migration with the alternative substrate (R)-1-HPP. This transformation probably involves an intermediary carbocation, based on observations with additional substrate analogues, such as (1R)-1-hydroxyl-2-aminopropylphosphonate, and model reactions for both radical- and carbocation-mediated migration. The ability of HppE to catalyse distinct reactions depending on the regio- and stereochemical properties of the substrate is given a structural basis using X-ray crystallography. These results provide compelling evidence for the formation of a substrate-derived cation intermediate in the catalytic cycle of a mononuclear non-haem-iron-dependent enzyme. The underlying chemistry of this unusual phosphono migration may represent a new paradigm for the *in vivo* construction of phosphonate-containing natural products that can be exploited for the preparation of new phosphonate derivatives.

Fosfomycin (1) is a clinically useful antibiotic for the treatment of limb-threatening diabetic foot infections and lower urinary tract infections<sup>4</sup>. Its biological target is the bacterial enzyme MurA, which catalyses the first committed step in the biosynthesis of peptidoglycan, the main component of the bacterial cell wall<sup>6,7</sup>. Notably, fosfomycin has been shown to be effective against ciprofloxacin-resistant *Escherichia coli*<sup>8</sup>, and methicillin- or vancomycin-resistant strains of *Staphylococcus aureus*<sup>9,10</sup>. Fosfomycin is derived from phosphoenolpyruvate (PEP, 2), with the phosphonate moiety being generated in a rearrangement reaction catalysed by PEP phosphomutase (Fom1) (2→3, Fig. 1a)<sup>11,12</sup>. This is followed by a reaction sequence involving decarboxylation, reduction and carbon methylation to produce (S)-2-HPP (4)<sup>13,14</sup>. The final step of the pathway is the conversion of 4 to 1, catalysed by HppE<sup>1,2</sup>.

Early studies showed that HppE is a mononuclear non-haem-iron-dependent enzyme<sup>1,2</sup> that uses a 2-His-1-carboxylate facial triad as the iron ligands, with its overall structural fold belonging to the cupin superfamily<sup>3,15</sup>. However, unlike many other enzymes in this class<sup>16</sup>, its activity is not dependent on  $\alpha$ -ketoglutarate. Instead, HppE uses reducing equivalents derived from NADH to activate molecular oxygen<sup>1,2</sup>. In addition, isotope-labelling experiments demonstrated that the oxygen atom of the oxiranyl ring in 1 is not derived from  $O_2$ , but instead from the secondary hydroxyl group of 4 (Fig. 1a)<sup>1,5</sup>. Thus, the HppE-catalysed conversion of 4 to 1 is in fact a dehydrogenation reaction and not an oxygenation reaction.

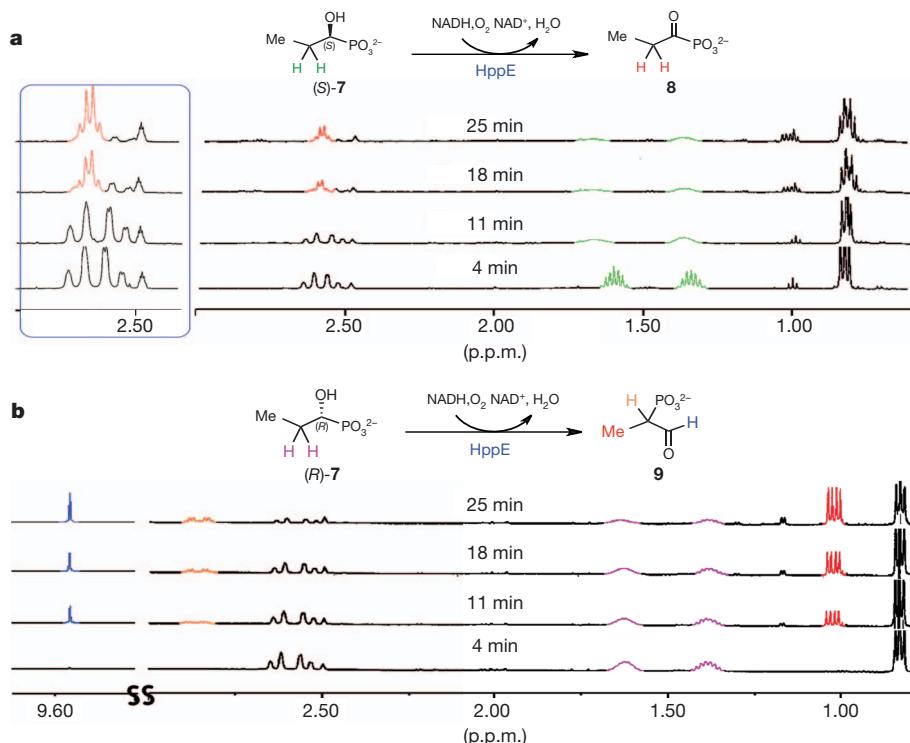
Previous experiments have also shown that the HppE reaction course is dependent on the substrate stereochemistry, as the HppE-catalysed dehydrogenation of (R)-2-HPP (5) produces the ketone 6 rather than an epoxide (Fig. 1b)<sup>17</sup>. Recent spectroscopic and crystallographic studies revealed that both enantiomers of 2-HPP (4 and 5) act as bidentate ligands to the mononuclear iron, such that only a single hydrogen atom is poised for abstraction by a reactive iron–oxygen species<sup>3,18,19</sup>. The direct coordination of the negatively charged phosphonate group to the iron centre probably helps to activate Fe(II) for reaction with  $O_2$ , and thus facilitates the formation of higher iron oxidation states for substrate oxidation<sup>3,20,21</sup>. These findings prompted a more thorough examination of the substrate flexibility and reactivity of HppE. Towards this aim, both enantiomers of 1-hydroxypropylphosphonate (7, 1-HPP) were synthesized and used as mechanistic probes of the HppE-catalysed reaction (Supplementary Information).



**Figure 1 | Fosfomycin biosynthetic pathway and HppE-catalysed conversion of various substrate analogues.** **a**, Formation of fosfomycin (1) from PEP (2). The oxygen atom of the hydroxyl group of 4 is shown in red to highlight its retention in the fosfomycin product. Fom1, PEP phosphomutase; Fom2, phosphonopyruvate decarboxylase; FomC, phosphonocetaldehyde dehydrogenase; Fom3, 2-hydroxyethylphosphonate C-methylase. **b**, Conversion of (R)-2-HPP (5) to the corresponding ketone 6. **c**, Conversion of (S)-1-HPP ((S)-7) to acyl phosphonate (8). **d**, Conversion of (R)-1-HPP ((R)-7) to the aldehyde product (9). The <sup>13</sup>C-labelled carbons are indicated in red.

<sup>1</sup>Division of Medicinal Chemistry, College of Pharmacy, and Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, USA. <sup>2</sup>Howard Hughes Medical Institute and the Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>†</sup>Present addresses: Department of Chemistry, Penn State University, University Park, Pennsylvania 16802, USA (W.-c.C.); Department of Chemistry, University of Iowa, Iowa City, Iowa 52242, USA (M.D.); Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA (P.L.); Department of Chemistry and Biochemistry, Auburn University, Auburn, Alabama 36849, USA (S.O.M.); Agensys, Santa Monica, California 90404, USA (S.-J.M.); Department of Biotechnology, Dalian Institute of Chemical Physics, Dalian 116023, China (Z.K.Z.).

\*These authors contributed equally to this work.

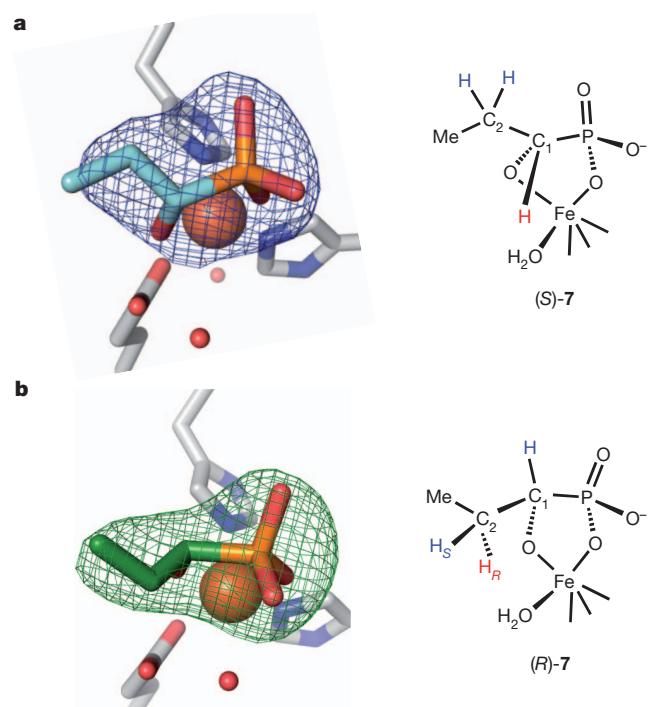


**Figure 2 |** <sup>1</sup>H NMR time-course for the HppE-catalysed conversion of (S)-7 to 8 (a) and (R)-7 to 9 (b). The peak at  $\delta$  2.49 is from deuterated dimethylsulfoxide (DMSO-d<sub>6</sub>), and those (in black) centred between  $\delta$  2.50

and 2.65 are from NADH (this region is enlarged in the box). The NMR signals and the contributing proton(s) are colour coded.

Incubation of (S)-7 with HppE produced the acyl phosphonate 8 (Figs 1c and 2a), a reaction analogous to the dehydrogenation of 5 to form the corresponding C2 ketone (6, Fig. 1b). Both reactions are consistent with hydrogen-atom abstraction from the oxygen-bearing carbon to yield an  $\alpha$ -hydroxyalkyl radical that undergoes one-electron oxidation to form the corresponding oxo product. By contrast, when (R)-7 was treated with HppE, 1-oxopropan-2-ylphosphonate (9) was obtained as the sole product (Figs 1d and 2b). The structure of 9 was initially determined by NMR, and was verified after NaBH<sub>4</sub> reduction to the more stable product 1-(hydroxymethyl)ethylphosphonate (10) and comparison with the chemically synthesized standard<sup>22</sup>. To further validate the structure of 9, (R)-7 was prepared enriched with <sup>13</sup>C at C1 ((R)-[1-<sup>13</sup>C]-7) and used as the substrate in the HppE reaction (Supplementary Information). Analysis by <sup>13</sup>C NMR demonstrated that the large C1-P coupling constant of 150.0 Hz in (R)-[1-<sup>13</sup>C]-7 was reduced to 2.7 Hz in 10 following reaction with HppE. These results indicate that HppE catalyses a 1,2-shift of the phosphono group during catalytic turnover of the substrate analogue (R)-7, for which there is, to our knowledge, no enzymatic precedent.

The crystal structures of Fe(II)-HppE in complex with (R)-7 and (S)-7 were determined to 2.71 and 2.80 Å resolution, respectively, in order to provide a structural basis for the altered stereo- and regiospecificity of the HppE-catalysed reactions with these substrates (Supplementary Information). Our results show that (R)-7 and (S)-7 bind to the iron centre in a bidentate fashion through an oxygen from the phosphonate moiety and the C1 hydroxyl (Fig. 3 and Supplementary Figs 5 and 6). Three additional coordination sites of the iron are filled by the 2-His-1-carboxylate facial triad, leaving one site available for dioxygen binding. This site, partially occupied by a water molecule in these structures, is the same putative O<sub>2</sub>-binding pocket to which the O<sub>2</sub>-mimic nitric oxide binds<sup>8</sup>. This spatial arrangement would direct the C1 hydrogen of (S)-7 and the *pro-R* C2 hydrogen of (R)-7 towards the reactive iron–oxygen species, consistent with the observed reaction products 8 and 9.



**Figure 3 |** Structures of (S)-7 and (R)-7 bound to the iron centre of HppE in a bidentate mode. a, (S)-7 (carbons in light blue) in  $F_o - F_c$  omit map density contoured at  $6\sigma$  (left) and its chemical structure (right), with hydrogen atoms accessible for abstraction in red and inaccessible in blue. The putative dioxygen binding site on Fe is partially occupied by water molecules (red spheres) in both of these structures (Fe to H<sub>2</sub>O distances are 3.7 and 3.0 Å). Colours: Fe in rust, P in orange, O in red, N in blue, C in grey. b, (R)-7 (carbons in green) in  $F_o - F_c$  omit map density contoured at  $6\sigma$  (left) and its chemical structure (right). Hydrogen atoms labelled as in a.

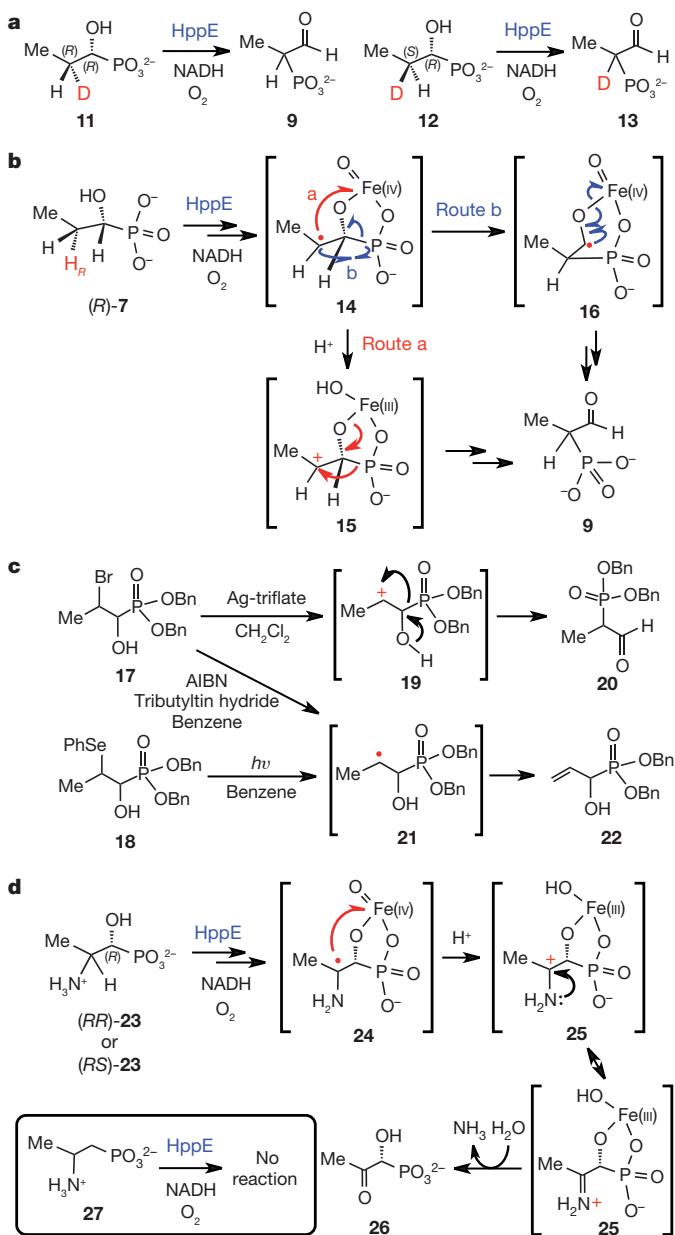
To verify the stereochemistry of hydrogen-atom abstraction from the prochiral carbon of (*R*)-7, substrate stereospecifically deuterated at the *pro-R* (11) or *pro-S* (12) positions of C2 were synthesized (Supplementary Fig. 2) and reacted with HppE. <sup>1</sup>H NMR analysis of the methyl-group splitting pattern of the reaction products from both diastereomers (Supplementary Fig. 3) demonstrated retention of only the *pro-S* hydrogen, consistent with the stereospecific abstraction of the *pro-R* hydrogen atom from C2 of (*R*)-7 (Fig. 4a and Supplementary Information) predicted by crystal structural analysis (Fig. 3).

Although non-enzymatic 1,2-phosphono migrations are known, they require either strong Lewis acids<sup>23</sup> or harsh alkaline conditions<sup>24</sup> and are generally thought to proceed through carbocationic intermediates. It is therefore plausible that such an intermediate may also be involved in the HppE-catalysed 1,2-phosphono migration of (*R*)-7 (Fig. 4b, route a). Following *pro-R* hydrogen-atom abstraction by a reactive iron–oxygen species to form the C2-centred radical 14, electron transfer to the iron centre would produce the corresponding C2 carbocation 15. This oxidation would then trigger the 1,2-phosphono migration in direct analogy to the non-enzymatic reactions. However, an alternative route involving a C2 radical-mediated migration of the C–P bond to generate the ketyl radical 16 before electron transfer could not be excluded (Fig. 4b, route b).

To gain insight into the chemistry of the defining intermediates of these two mechanistic hypotheses, model studies were carried out with compounds 17 and 18 (Supplementary Information). As summarized in Fig. 4c, when 17 was treated with silver triflate to generate the carbocation intermediate 19, the migration product 20 was produced. By contrast, when the alkyl radical 21 was generated, either by exposure of compound 18 to ultraviolet radiation or treatment of 17 with tributyltin hydride, formation of alkene 22 was observed. These findings provide a correlation between formation of a carbocation intermediate and 1,2-phosphono migration. It is also worth noting that the intermediacy of carbocationic species has been implicated in most biological 1,2-hydride and 1,2-alkyl shift reactions, such as those catalysed by various terpenoid cyclases<sup>25</sup> and the ‘NIH shift’ of aromatic amino acid hydroxylases<sup>26</sup>. However, the situation could still be different on the enzyme, as in the model reaction formation of 22 from 21 may simply be more rapid than rearrangement to form 20 in solution.

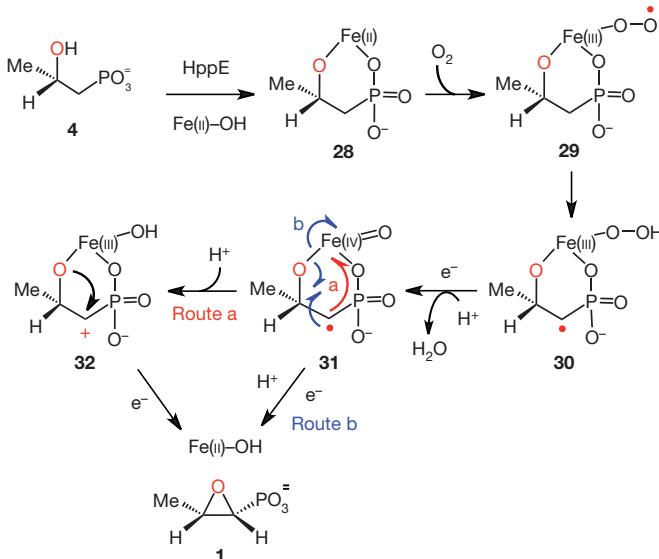
Further evidence supporting a carbocationic intermediate in the HppE-catalysed reaction was obtained using the substrate analogues (1*R*, 2*R*)- and (1*R*, 2*S*)-1-hydroxyl-2-aminopropyl phosphonate ((*RR*)- and (*RS*)-23; Fig. 4d and Supplementary Information). Using a NMR assay and authentic standards (Supplementary Fig. 4 and Supplementary Information), HppE was shown to convert both isomers of 23 to 26 with no detectable formation of the migration or other product. This ketone product can be envisioned as forming through hydrogen-atom abstraction from C2, yielding an  $\alpha$ -aminoalkyl radical (24) that is oxidized to the stable C2 iminium ion 25, which can undergo spontaneous hydrolysis to yield 26. Hydrogen-atom abstraction from C2 and the conversion of both isomers of 23 to 26 are consistent with bidentate coordination to the active site iron through the C1 hydroxyl and phosphonate moieties, in a manner analogous to (*R*)-7 (Supplementary Figs 7 and 8). Support for the hypothesis that the C2 amino group of 23 is unable to sustain the bidentate substrate coordination required for catalysis was obtained with the substrate analogue 2-aminopropylphosphonate (( $\pm$ )-27; Fig. 4d), which lacks the 1-hydroxyl group of 23 and is not a substrate of HppE (Supplementary Information). Taken together, these results are consistent with a mechanism in which a reactive iron–oxygen species generated during the catalytic cycle of HppE is capable of oxidizing the C2-centred radical 24 to the corresponding carbocation 25 during the conversion of 23 to 26 (Fig. 4d).

In summary, this investigation provides new insight into the catalytic capability and chemical mechanism of the non-haem-iron enzyme HppE. HppE shows remarkable catalytic versatility, converting the (*R*) and (*S*) isomers of 1-HPP and 2-HPP to aldehyde, acyl phosphonate,



**Figure 4 | Summary of HppE-catalysed reactions and model chemistry examined in this study.** **a**, Stereochemistry of hydrogen-atom abstraction from (*R*)-1-HPP (7) determined using the stereospecifically deuterated compounds 11 and 12. The deuterium atom is highlighted in red. **b**, Hypothetical mechanisms for HppE-catalysed 1,2-phosphono migration involving cation (route a, red arrows)-mediated and radical (route b, blue arrows)-mediated rearrangements. Intermediate 14 could be generated in a manner analogous to the formation of 31 from (*S*)-2-HPP (see Fig. 5) via abstraction of the *pro-R* C2 hydrogen (highlighted in red). **c**, Model reactions to probe the involvement of radical or cation intermediates in the 1,2-phosphono migration catalysed by HppE. The migration product is only observed when cation 19 is formed from 17 using silver triflate. **d**, HppE-catalysed conversion of (*RR*)-23 and (*RS*)-23 to the imine hydrolysis product 26, consistent with the oxidation of (*R*)-7 to a cationic intermediate by a reactive iron–oxygen species (red arrow). Box, compound 27 (which lacks the C1 hydroxyl group) is not a substrate of HppE, indicating that the amino group is not capable of supporting the bidentate substrate coordination required for catalysis.

ketone and epoxide products, respectively (Fig. 1). This study reveals an unprecedented 1,2-phosphono-migration reaction and provides support for the existence of carbocation intermediates in the HppE reaction. The unique chemistry observed during the phosphono-migration reaction catalysed by HppE may represent a new paradigm



**Figure 5 |** Possible mechanisms for the HppE-catalysed epoxidation of (S)-2-HPP (4) involving C1 cation formation (route a, red arrow) or oxygen-atom rebound (route b, blue arrows). The oxygen atom of the hydroxyl group of 4 is shown in red to highlight its retention in the fosfomycin product.

for the rearrangement of C–P bonds in nature, which can be exploited for the synthesis of new phosphonate-containing natural products. Moreover, the probable use of a carbocation intermediate in the HppE-catalysed conversion of 23 to 26 buoys the proposal of a carbocation (15) intermediate in the conversion of (R)-7 to 9 (phosphonomigration product in Fig. 1d), and prompts re-examination of the mechanism of HppE in general. In terms of the types of iron–oxygen species used by HppE,  $^{18}\text{O}$  kinetic isotope effect studies with the natural substrate, (S)-2-HPP (4), demonstrate partially rate-limiting formation of a Fe(III)–hydroperoxy intermediate (30)<sup>27</sup>, which implicates Fe(III)–superoxide as the species that probably abstracts the hydrogen atom from the substrate (29→30; Fig. 5)<sup>28,29</sup>. At first glance, the C1–H bond of (S)-2-HPP (29 in Fig. 5) might seem more activated than the C2–H bond of (R)-7 (Fig. 4), with the latter bond cleavage requiring the use of a highly reactive Fe(IV)–oxo species. However, density functional theory calculations indicate that the phosphonate moiety of the substrate provides significant  $\beta$ -stabilization to the C2-centred radical (14), such that the bond dissociation energy of the C2–H of (R)-1-HPP is actually  $\sim 2.7\text{ kcal mol}^{-1}$  less than that of the C1–H of (S)-2-HPP (93.8 versus 96.5  $\text{kcal mol}^{-1}$ ) (Supplementary Information). Thus, a Fe(III)–superoxide species may also be used by HppE to effect hydrogen-atom abstraction from C2 of (R)-7, allowing the Fe(IV)–oxo intermediate generated in a subsequent step to be used for the oxidation of the C2 radical to the cationic intermediate (14→15, Fig. 4b). This chemical logic could apply to HppE in its reaction with the natural substrate. With a Fe(III)–superoxide intermediate responsible for hydrogen-atom abstraction, an Fe(IV)–oxo intermediate is available to catalyse the more challenging oxidation of a C1 radical to a carbocation (31→32, Fig. 5). Indeed, this may be a more common strategy than previously surmised, being used by the growing number of identified non-haem-iron-dependent enzymes that initiate substrate oxidation using iron-superoxide intermediates<sup>29</sup>. The formation of the Fe(IV)–oxo intermediate 31 requires the input of a single electron at the stage of a highly reactive substrate radical intermediate (30). The precise timing of this redox reaction may be accomplished by the transfer of an electron from the putative proton-coupled electron-transfer pathway of HppE, which comprises several tyrosine residues<sup>3</sup>. The resulting protein radical could then be quenched by the external reductant at a subsequent (and not necessarily precisely controlled) time to regenerate the active form of the enzyme. Experiments to characterize the catalytic mechanism of HppE further are in progress.

## METHODS SUMMARY

HppE used in this study was purified as described previously<sup>28</sup>. All synthetic and enzymatic reaction products were characterized by NMR spectroscopy and/or high-resolution mass spectrometry. A typical NMR assay contained 0.25 mM HppE, 0.25 mM  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ , 7.5 mM flavin mononucleotide (FMN), 25 mM substrate ((R)-7, (S)-7, 11, 12, (RR)-23, (RS)-23 or 27) and 25 mM NADH in 700  $\mu\text{l}$  20 mM Tris buffer (pH 7.5). The conversion of (RR)-23/(RS)-23 to 26 by HppE was confirmed by spiking the reaction mixture with an authentic standard of 26. Further details regarding experimental procedures and density functional theory calculations are described in Supplementary Information.

Received 5 April 2012; accepted 8 February 2013.

- Liu, P. et al. Protein purification and function assignment of the epoxidase catalyzing the formation of fosfomycin. *J. Am. Chem. Soc.* **123**, 4619–4620 (2001).
- Liu, P. et al. Biochemical and spectroscopic studies on (S)-2-hydroxypropylphosphonic acid epoxidase: a novel mononuclear non-heme iron enzyme. *Biochemistry* **42**, 11577–11586 (2003).
- Higgins, L. J., Yan, F., Liu, P., Liu, H.-w. & Drennan, C. L. Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. *Nature* **437**, 838–844 (2005).
- Stengel, D. et al. Second-line treatment of limb-threatening diabetic foot infections with intravenous fosfomycin. *J. Chemother.* **17**, 527–535 (2005).
- Hammerschmidt, F., Boermann, G. & Bayer, K. Biosynthesis of natural products with a P–C Bond. 5. The oxirane oxygen atom of fosfomycin is not derived from atmospheric oxygen. *Justus Liebigs Ann. Chem.* 1055–1061 (1990).
- Marquardt, J. L. et al. Kinetics, stoichiometry, and identification of the reactive thiolate in the inactivation of UDP-GlcNAc enolpyruvyl transferase by the antibiotic fosfomycin. *Biochemistry* **33**, 10646–10651 (1994).
- Brown, E. D., Vivas, E. I., Walsh, C. T. & Kolter, R. MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*. *J. Bacteriol.* **177**, 4194–4197 (1995).
- Ko, K. S. et al. *In vitro* activity of fosfomycin against ciprofloxacin-resistant or extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* isolated from urine and blood. *Diagn. Microbiol. Infect. Dis.* **58**, 111–115 (2007).
- Nakazawa, H., Kikuchi, Y., Honda, T., Isago, T. & Nozaki, M. Enhancement of antimicrobial effects of various antibiotics against methicillin-resistant *Staphylococcus aureus* (MRSA) by combination with fosfomycin. *J. Infect. Chemother.* **9**, 304–309 (2003).
- Cassone, M., Campanile, F., Pantosti, A., Venditti, M. & Stefanini, S. Identification of a variant ‘Rome clone’ of methicillin-resistant *Staphylococcus aureus* with decreased susceptibility to vancomycin, responsible for an outbreak in an intensive care unit. *Microb. Drug Resist.* **10**, 43–49 (2004).
- Seidel, H. M., Freeman, S., Seto, H. & Knowles, J. R. Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon–phosphorus bond. *Nature* **335**, 457–458 (1988).
- Liu, S., Lu, Z., Jia, Y., Dunaway-Mariano, D. & Herzberg, O. Dissociative phosphoryl transfer in PEP mutase catalysis: structure of the enzyme/sulfonylpyruvate complex and kinetic properties of mutants. *Biochemistry* **41**, 10270–10276 (2002).
- Hidaka, T. et al. Cloning and nucleotide sequence of fosfomycin biosynthetic genes of *Streptomyces wedmorensis*. *Mol. Gen. Genet.* **249**, 274–280 (1995).
- Woodyer, R. D., Li, G. Y., Zhao, H. M. & van der Donk, W. A. New insight into the mechanism of methyl transfer during the biosynthesis of fosfomycin. *Chem. Commun.* 359–361 (2007).
- Yan, F., Li, T., Lipscomb, J. D., Liu, A. & Liu, H.-w. Site-directed mutagenesis and spectroscopic studies of the iron-binding site of (S)-2-hydroxypropylphosphonic acid epoxidase. *Arch. Biochem. Biophys.* **442**, 82–91 (2005).
- Kovaleva, E. G. & Lipscomb, J. D. Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nature Chem. Biol.* **4**, 186–193 (2008).
- Zhao, Z. et al. Mechanistic studies of HPP epoxidase: configuration of the substrate governs its enzymatic fate. *Angew. Chem. Int. Edn Engl.* **41**, 4529–4532 (2002).
- Yun, D. et al. Structural basis of regiospecificity of a mononuclear iron enzyme in antibiotic fosfomycin biosynthesis. *J. Am. Chem. Soc.* **133**, 11262–11269 (2011).
- Yan, F. et al. Determination of the substrate binding mode to the active site iron of (S)-2-hydroxypropylphosphonic acid epoxidase using  $^{17}\text{O}$ -enriched substrates and substrate analogues. *Biochemistry* **46**, 12628–12638 (2007).
- Costas, M., Mehn, M. P., Jensen, M. P. & Que, L., Jr. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* **104**, 939–986 (2004).
- Huang, H. et al. Evidence for radical-mediated catalysis by HppE: a study using cyclopropyl and methylenecyclopropyl substrate analogues. *J. Am. Chem. Soc.* **134**, 13946–13949 (2012).
- Yamashita, M., Sugiyama, M., Oshikawa, T. & Inokawa, S. Synthesis of dimethyl (1-nitromethylalkyl)phosphonates and their conversion to dimethyl (1-formylalkyl)phosphonates by oxidation with ozone. *Synthesis* **1987**, 62–64 (1987).
- Churi, R. H. & Griffin, C. E. 1,2 shifts of dialkoxyphosphono groups in skeletal rearrangements of  $\alpha,\beta$ -epoxyvinylphosphonates. *J. Am. Chem. Soc.* **88**, 1824–1825 (1966).
- Girotra, N. N. & Wendler, N. L. Synthesis and transformations in the phosphonomycin series. *Tetrahedr. Lett.* **10**, 4647–4650 (1969).

25. Lesburg, C. A., Caruthers, J. M., Paschall, C. M. & Christianson, D. W. Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism. *Curr. Opin. Struct. Biol.* **8**, 695–703 (1998).
26. Kappock, T. J. & Caradonna, J. P. Pterin-dependent amino acid hydroxylases. *Chem. Rev.* **96**, 2659–2756 (1996).
27. Mirica, L. M., McCusker, K. P., Munos, J. W., Liu, H.-w. & Klinman, J. P.  $^{18}\text{O}$  kinetic isotope effects in non-heme iron enzymes: probing the nature of Fe/O<sub>2</sub> intermediates. *J. Am. Chem. Soc.* **130**, 8122–8123 (2008).
28. Yan, F., Munos, J. W., Liu, P. & Liu, H.-w. Biosynthesis of fosfomycin, re-examination and re-confirmation of a unique Fe(II)- and NAD(P)H-dependent epoxidation reaction. *Biochemistry* **45**, 11473–11481 (2006).
29. van der Donk, W. A., Krebs, C. & Bollinger, J. M. Substrate activation by iron superoxo intermediates. *Curr. Opin. Struct. Biol.* **20**, 673–683 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank W. Johnson and S. Sorey for assistance with the NMR experiment setup. This work is supported in part by grants from The National Institutes of Health (GM040541 to H.-w.L.) and the Welch Foundation (F-1511). C.L.D. is a

Howard Hughes Medical Institute Investigator. Crystallographic data collection were conducted at the Advanced Light Source, a Department of Energy (DOE) national user facility (Contract DE-AC02-05CH11231), at beamline 8.2.2 operated by the Berkeley Center for Structural Biology, which is supported in part by DOE and National Institutes of Health/National Institute of General Medical Sciences.

**Author Contributions** H.-w.L. provided the scientific direction and overall experimental design for the studies. W.-c.C. performed many of the experiments described herein. M.D. and C.L.D. collected and interpreted all of the crystallographic data. P.L., S.-J.M. and Z.K.Z. designed and carried out the initial biochemical as well as model system studies. S.O.M. performed the DFT calculations. W.-c.C., M.D., P.L., S.O.M., C.L.D. and H.-w.L. wrote the manuscript.

**Author Information** Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 4J1W and 4J1X. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-w.L. ([h.w.liu@mail.utexas.edu](mailto:h.w.liu@mail.utexas.edu)).

# A solution to release twisted DNA during chromosome replication by coupled DNA polymerases

Isabel Kurth<sup>1</sup>, Roxana E. Georgescu<sup>1</sup> & Mike E. O'Donnell<sup>1</sup>

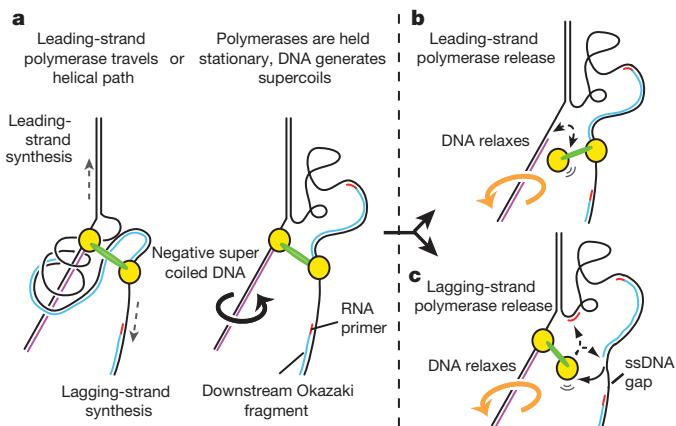
Chromosomal replication machines contain coupled DNA polymerases that simultaneously replicate the leading and lagging strands<sup>1</sup>. However, coupled replication presents a largely unrecognized topological problem. Because DNA polymerase must travel a helical path during synthesis, the physical connection between leading- and lagging-strand polymerases causes the daughter strands to entwine, or produces extensive build-up of negative supercoils in the newly synthesized DNA<sup>2–4</sup>. How DNA polymerases maintain their connection during coupled replication despite these topological challenges is unknown. Here we examine the dynamics of the *Escherichia coli* replisome, using ensemble and single-molecule methods, and show that the replisome may solve the topological problem independent of topoisomerases. We find that the lagging-strand polymerase frequently releases from an Okazaki fragment before completion, leaving single-strand gaps behind. Dissociation of the polymerase does not result in loss from the replisome because of its contact with the leading-strand polymerase. This behaviour, referred to as ‘signal release’, had been thought to require a protein, possibly primase, to pry polymerase from incompletely extended DNA fragments<sup>5–7</sup>. However, we observe that signal release is independent of primase and does not seem to require a protein trigger at all. Instead, the lagging-strand polymerase is simply less processive in the context of a replisome. Interestingly, when the lagging-strand polymerase is supplied with primed DNA *in trans*, uncoupling it from the fork, high processivity is restored. Hence, we propose that coupled polymerases introduce topological changes, possibly by accumulation of superhelical tension in the newly synthesized DNA, that cause lower processivity and transient lagging-strand polymerase dissociation from DNA.

Coupled leading- and lagging-strand polymerases, as observed in phage and bacterial replisomes<sup>1,8</sup>, have marked implications for the topology of the daughter DNA products<sup>2–4,9</sup>. The topological problem is based on the *a priori* fact that polymerases generate a helical product, and thus either they must travel a helical path or the DNA product must turn behind them. For example, a rotating leading-strand polymerase will take the attached lagging-strand polymerase and wind the DNA duplex 360° around the axis of the leading strand, forming precatenanes in the daughter helices (Fig. 1a, left)<sup>2</sup>. At the rate of *E. coli* replication (650 base pairs per second) rotating polymerases quickly result in an impossible tangle<sup>2</sup>. Alternatively, if the DNA products rotate instead of the polymerases, negative supercoils accumulate (Fig. 1a, right). Superhelical tension on the lagging strand may be relieved by rotation of single-stranded DNA (ssDNA), but ssDNA-binding (SSB) proteins form large superstructures, which probably constrain swivel motion<sup>10</sup>. The energy generated by only three or four supercoils<sup>11</sup> is sufficient to disrupt protein–protein and protein–DNA interactions<sup>12</sup>. Thus topological tension could disrupt the replisome, unless the tension is periodically released. The supercoils produced by coupled polymerases behind the fork are negative supercoils, which can be resolved by Topo I and Topo III<sup>13</sup>. However, Topo III is non-essential, and the viability of Topo I mutants has been controversial<sup>14,15</sup>. Hence,

topoisomerases may participate, but are insufficient to remove negative supercoils produced by coupled replisomes.

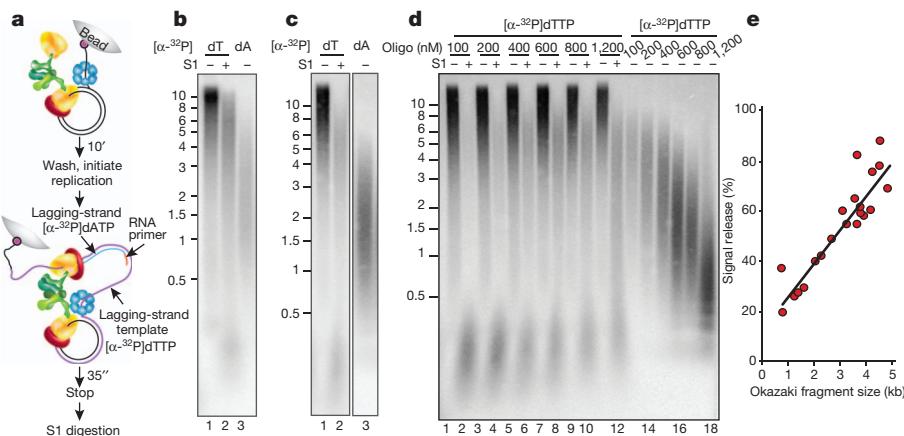
Interestingly, single-molecule studies demonstrate highly processive DNA synthesis (>100 kilobases (kb)) in the absence of topoisomerases, implying that replisomes possess an intrinsic solution to the problem of coupled replication<sup>16,17</sup>. Indeed, it was originally proposed that the topological problem could be solved by transient release of one polymerase of a coupled replisome from DNA (Fig. 1b), enabling the negative supercoils on both strands to relax<sup>2</sup>. If the leading polymerase detaches from DNA, it will rebind the same primer terminus for continued extension. If the lagging polymerase dissociates, it will reattach to the same Okazaki fragment and complete it provided a new primed site is not yet formed. During extension of longer Okazaki fragments a new priming event is more likely to occur; thus a dissociated lagging polymerase may rebind the new RNA primer, leaving the original Okazaki fragment incomplete (Fig. 1b). In fact incomplete Okazaki fragments have been observed in bacterial and phage systems<sup>5–7</sup>. Premature polymerase dissociation is referred to as ‘signal release’ because it has been presumed that the lagging-strand polymerase is ‘signalled’ by a replisome component to dissociate and alleviate supercoil tension<sup>5–7</sup>.

To gain insight into the topological challenge of coupled DNA replication, we developed assays to study signal release. The *E. coli* replisome is assembled on a 5' biotinylated rolling circle substrate and then attached to streptavidin beads (Fig. 2a). The substrate has



**Figure 1 | The topological problem caused by coupled leading- and lagging-strand polymerases.** The figure illustrates the topological problem only for the leading-strand polymerase. **a**, The two Pol (yellow) are coupled through the clamp loader (green). Pol travel a helical path as they synthesize DNA. As the leading Pol spins, the lagging Pol is dragged around this path, twisting the lagging strand around the leading strand (left). Alternatively, the DNA turns instead (right), generating negative supercoils as indicated by the arrow. In either case, the stress can be released by **(b)** transient dissociation of the leading Pol or **(c)** transient dissociation of the lagging Pol, which could re-bind the incomplete Okazaki fragment or a new RNA primer.

<sup>1</sup>The Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA.



**Figure 2 | Signal release does not require primase and correlates with increasing Okazaki fragment length.** **a**, The replisome, consisting of DNA helicase (blue), Pol III\* (one clamp loader (green) that binds three Pols (yellow); only two Pols are shown for clarity) and  $\beta$ -clamps (red) is assembled on a 5' biotinylated rolling circle DNA, then attached to beads. After unbound proteins are removed, replication is initiated by adding primase,  $\beta$ , SSB, ATP and  $\alpha$ -<sup>32</sup>P-labelled deoxyribonucleotide triphosphates. Reactions are quenched and treated with S1 to cleave gaps left by signal release, before analysis on

only three nucleotides on either strand, allowing either leading or lagging strand to be labelled depending on which radioactive nucleotide is used. Replication is initiated after wash steps to remove unbound proteins, and ssDNA gaps left by signal release are detected by treating the  $\alpha$ -<sup>32</sup>P-labelled deoxythymidine triphosphate ([ $\alpha$ -<sup>32</sup>P]dTTP)-labelled lagging-strand template with S1 endonuclease to digest the gaps (Supplementary Fig. 1). If all Okazaki fragments are completely extended, the product will be S1-resistant. Conversely, 100% signal release will leave gaps between every Okazaki fragment, and S1 digestion will yield products of similar size to Okazaki fragments (Supplementary Fig. 2). An intermediate length of DNA will result if signal release is less than 100%.

The S1 analysis shows an intermediate-sized [ $\alpha$ -<sup>32</sup>P]dTTP-labelled DNA (Fig. 2b, lane 2) that migrates between the length of undigested DNA (lane 1) and the size of Okazaki fragments (lane 3). Size analysis of S1-treated products shows that about one-third of all Okazaki fragments are incomplete (Supplementary Fig. 3 and Methods). To validate the assay, several controls were performed (Supplementary Fig. 4): (1) S1 digestion eliminates ssDNA produced in a reaction without primase (that is, no lagging-strand synthesis); (2) S1 does not cleave double-strand Okazaki fragments; (3) ssDNA gaps produced by signal release can be filled in by addition of DNA polymerase II (Pol II); (4) the presence of complete Okazaki fragments is confirmed using ligase.

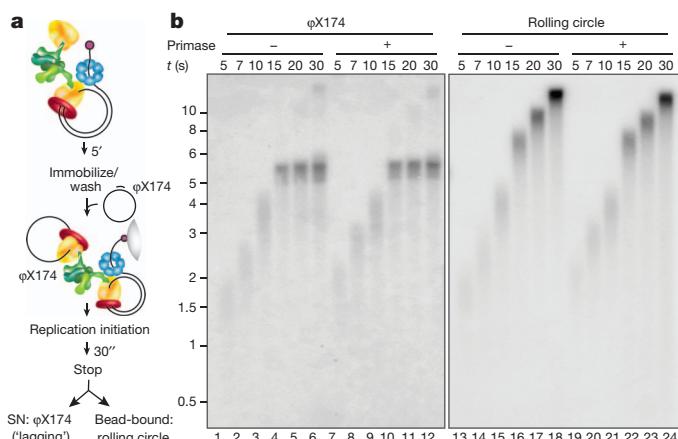
Earlier studies implied that primase triggers polymerase signal release<sup>5,6</sup>. To test this hypothesis, we performed the reaction in the complete absence of primase, using exogenously added DNA primers to initiate Okazaki fragments. However, the result was unchanged (Fig. 2c). Studies in the T4 system suggest that a loaded clamp triggers polymerase release<sup>7</sup>. Although clamps cannot be omitted without negating lagging-strand synthesis altogether, titrations of the clamp (with or without clamp loader) do not influence signal release (Supplementary Fig. 5). Notably, the assays of Fig. 2 remove excess (unbound) polymerase-clamp loader (that is, Pol III\*) and DnaB helicase, and show that primase, soluble polymerase, clamp loader and helicase are not required to trigger polymerase dissociation.

We find that increasing Okazaki fragment size by lowering DNA primer concentration results in more frequent signal release (Fig. 2d, e and Supplementary Fig. 6). Okazaki fragments of 1 kb give only 20% signal release whereas 5 kb Okazaki fragments give 80% signal release. Hence, small Okazaki fragments are more likely to be completed than long Okazaki fragments, implying that the lagging-strand polymerase

alkaline gels. The newly synthesized leading strand (purple) is the template for lagging-strand synthesis (blue), initiated by RNA primers (red). **b**, Replication reactions using [ $\alpha$ -<sup>32</sup>P]dTTP either before (lane 1) or after (lane 2) S1 analysis, or using [ $\alpha$ -<sup>32</sup>P]dATP (lane 3). **c**, Reactions in the absence of primase, using 800 nM DNA 20-base oligonucleotides to prime the lagging strand. **d**, Titration of DNA 20-base oligonucleotides into reactions. **e**, Plot of Okazaki fragment length versus percentage signal release.

is not very processive. In contrast, studies on model lagging-strand templates show the polymerase has high intrinsic processivity and extends a primed 5.4 kb  $\phi$ X174 ssDNA without dissociating<sup>18,19</sup>. To exclude the possibility that this result was an artefact of using DNA primers, we altered Okazaki fragment size by changing the concentration of primase<sup>20</sup>, but the results were upheld (Supplementary Fig. 7). That signal release correlates with long Okazaki fragments was confirmed using Pol II to fill in ssDNA gaps, demonstrating that ssDNA gaps are associated with long Okazaki fragments and yield an average gap size of 800 nucleotides (Supplementary Fig. 8).

To test further for a protein circuit that lowers processivity by triggering polymerase release, we designed an assay to measure lagging-strand polymerase processivity directly. Because priming at a replication fork occurs randomly, producing heterogeneous-sized products, we attached a 5.4 kb primed  $\phi$ X174 circular ssDNA to the lagging-strand polymerase *in trans* before initiating replication (Fig. 3a). Considering that all replisomal proteins were present, protein-induced signal release should still have occurred, and the length of



**Figure 3 | Lagging-strand polymerase processivity is restored using a separate DNA molecule.** **a**, The replisome is assembled on DNA as in the legend to Fig. 2a. Unbound proteins are washed away before addition of primed  $\phi$ X174 5.4 kb ssDNA. **b**, Reactions are initiated in the presence (lanes 7–12 and 19–24) or absence (lanes 1–6 and 13–18) of primase. Supernatants containing  $\phi$ X174 replication products (lanes 1–12) are separated from bead-bound rolling circle products (lanes 13–24).

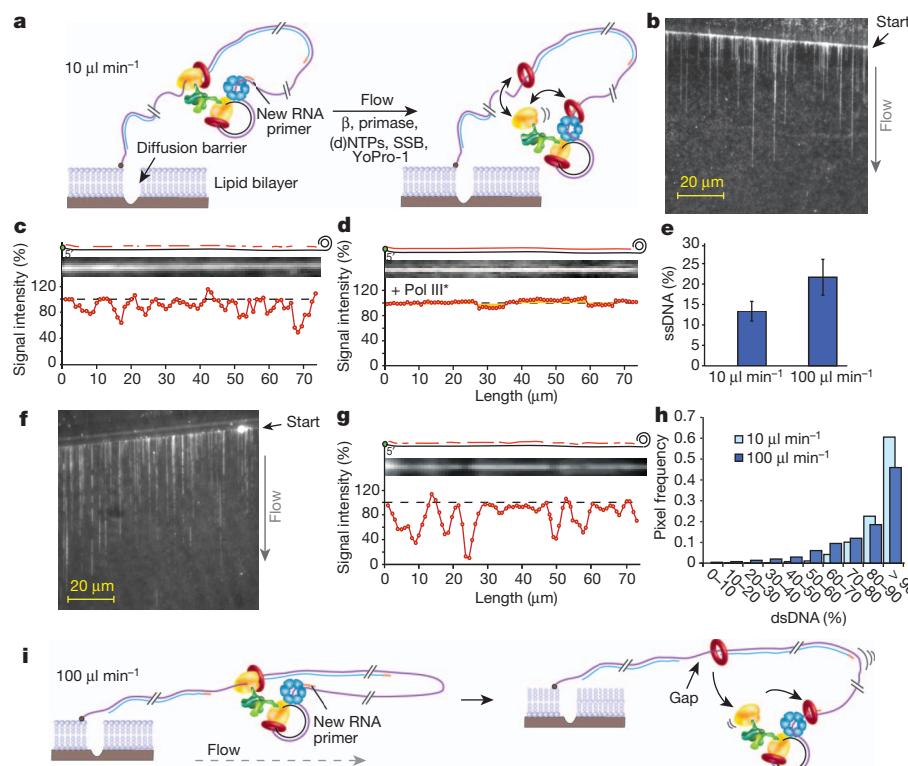
extension from the defined primed site would define the processivity of the lagging-strand polymerase.

A time-course of replication showed a most surprising result (Fig. 3b). The 5.4 kb  $\phi$ X174 DNA was fully replicated by the lagging polymerase (Fig. 3b, lanes 1–6 and 13–18), in striking contrast to the low processivity during Okazaki fragment synthesis (that is, 80% dissociation by 5 kb; Fig. 2d, e). We obtained similar results in the presence or absence of primase, confirming that neither primase nor primed sites trigger signal release (Fig. 3b, lanes 7–12 and 19–24). The main difference in reactions with or without the primed  $\phi$ X174 DNA *in trans* is the connection of the two polymerases through the DNA, indicating that the connection of two polymerases to the same DNA molecule causes a decrease in lagging-strand polymerase processivity. One possible non-protein source is the topological strain produced in DNA owing to coupled replication (Fig. 1). Lagging-strand loops are an unlikely source, as bending of large DNA stretches requires very little energy<sup>21</sup>. To examine the dynamics of lagging-strand polymerase action, we turned to single-molecule studies to determine whether signal release is characteristic of all replisomes, and whether it takes place during long replication times. Replication is performed in a flow cell in which the replisome–DNA complex is attached to a lipid bilayer. Replicating DNA is held in place at a lipid diffusion barrier (Fig. 4a and Methods). Upon initiating replication, product DNA is observed using total internal reflection fluorescence microscopy and a fluorescent double-stranded DNA (dsDNA) intercalator (see Fig. 4b). We initially performed experiments at  $10 \mu\text{l min}^{-1}$ , which translates into an applied force of 0.23 pN on the DNA, far below that needed to disrupt most protein–protein and protein–DNA interactions<sup>12</sup>. After replication was complete, we increased the flow to  $100 \mu\text{l min}^{-1}$  (1.45 pN) to stretch DNA, enabling examination of the product for ssDNA gaps, which appeared as regions with decreased

fluorescence intensity<sup>16</sup> (Fig. 4b, c). The resolution of the CCD (charge-coupled device) camera was approximately  $0.27 \mu\text{m}$  per pixel, the length of about 880 base pairs of dsDNA. Hence ssDNA gaps should have resulted in an uneven pixel density over the length of the molecule and facilitated a quantitative evaluation of the ssDNA/dsDNA ratio per pixel (Methods). All the molecules examined contained pixel intensity differences along their entire length (Fig. 4c and Supplementary Fig. 9). If the pixel intensity differences were indeed ssDNA, they should have been eliminated upon adding Pol III, which fills gaps, in the flow buffer, which was the case (Fig. 4d). Summation of the pixel intensities over  $298.4 \mu\text{m}$  of DNA in nine different DNA molecules indicated 13% ssDNA and 87% dsDNA (assuming SSB-bound ssDNA equals dsDNA length<sup>5</sup> (Fig. 4e)). This result is within twofold that observed in ensemble studies (that is, Okazaki fragments average 2.3 kb, and give 65% signal release at  $10 \text{nM}$  primase; thus an 800-base-pair average gap size yields a value of 22.3% ssDNA and 77.7% dsDNA). Overall, these results demonstrate that signal release occurs over the length of long DNA products and is a characteristic of all replisomes.

We also performed replication at a tenfold higher flow rate ( $100 \mu\text{l min}^{-1}$ ; 1.45 pN) (Fig. 4f, g). Unexpectedly, the amount of ssDNA relative to dsDNA increased nearly twofold relative to the  $10 \mu\text{l min}^{-1}$  flow rate (Fig. 4h), from 13% at  $10 \mu\text{l min}^{-1}$  to 22% at  $100 \mu\text{l min}^{-1}$  (Fig. 4e). An explanation for this difference is that when the lagging-strand polymerase undergoes a transient release cycle (signal release), the increased velocity of the flow carries away the ‘old’ 3’ terminus, making the polymerase more likely to associate with the next RNA primer synthesized at the fork (illustration in Fig. 4i) and resulting in more ssDNA gaps.

Notably, previous studies of signal release that suggested a protein trigger are also consistent with polymerase release triggered by torsional constraints owing to coupled replication<sup>5–7</sup>. For example, in T4,



**Figure 4 | Single-molecule total internal reflection fluorescence microscopy.** **a**, The replisome is assembled on the rolling circle DNA and attached to a lipid bilayer. Replication is initiated upon flowing replication buffer at  $10 \mu\text{l min}^{-1}$ . **b**, DNA curtain produced at  $10 \mu\text{l min}^{-1}$ . **c**, Normalized line plot analysis of a representative DNA strand. Diagrams at the top illustrate dsDNA and ssDNA regions. **d**, Analysis as in **c**, but with Pol III\* in the buffer

flow. **e**, Quantification of line plot analysis at indicated flow rates ( $n = 9 \pm \text{s.d.}$  for  $10 \mu\text{l min}^{-1}$  or  $n = 13 \pm \text{s.d.}$  for  $100 \mu\text{l min}^{-1}$ ). **f**, DNA curtain at  $100 \mu\text{l min}^{-1}$ . **g**, Line plot analysis as in **c**, at  $100 \mu\text{l min}^{-1}$ . **h**, Histogram relating pixel intensity to dsDNA at  $10 \mu\text{l min}^{-1}$  (light blue) and  $100 \mu\text{l min}^{-1}$  (dark blue). **i**, Scheme at  $100 \mu\text{l min}^{-1}$ .

increasing the concentration of clamps may accelerate clamp assembly on new RNA primers and increase the probability that a released polymerase binds the new RNA primed site<sup>7</sup>. In the T7 system, additional primase may enhance priming frequency providing more opportunity for a dissociated polymerase to bind a new site<sup>5</sup>.

Negative supercoiling is required for a variety of DNA metabolic actions, including replication, recombination and transcription<sup>22,23</sup>. Negative superhelicity is a property of genomic DNA in all organisms and seems constant over the entire genome<sup>24</sup>. Hence it is conceivable that coupled leading-/lagging-strand replication provides a method by which negative supercoils can be placed into DNA across the entire genome without topoisomerase action.

Our data suggest that signal release is caused by DNA topology and thus may provide a topoisomerase-independent solution to the topological problem incurred by coupled DNA polymerases during replication. The results demonstrate that (1) signal release does not require primase, (2) the release process is stochastic implying an inherent low processivity of the lagging polymerase, (3) signal release requires connection of both polymerases to the same DNA and (4) addition of DNA *in trans* eliminates signal release in the presence of all replisomal proteins. Despite the correlation of premature signal release to the topological problem inherent in coupled polymerases, the problem of coupled polymerase action is not definitively solved in this paper and will require further studies.

## METHODS SUMMARY

Reactions contained 100 fmol 5'-biotinylated 100-base oligonucleotide rolling circle DNA<sup>25</sup>, 60 μM deoxyadenosine triphosphate (dATP) and deoxyguanosine triphosphate (dGTP), 50 μM ATP-γ-S, 4 pmol DnaB<sub>6</sub>, 2.5 pmol β<sub>2</sub>, and 0.5 pmol Pol III\* in 20 μl replication buffer. Reactions were incubated 5 min at 37 °C before immobilization on streptavidin-magnetic beads. Beads were washed three times in 500 μl replication buffer containing 30 μM dATP, dGTP, 12 pmol β<sub>2</sub> and 50 μM ATP-γ-S before re-suspending in replication buffer containing 60 μM dCTP, 60 μM dGTP and 2.5 pmol β<sub>2</sub>. Replication was initiated by adding 0.5 mM ATP, 480 nM SSB<sub>4</sub>, 50 μM NTPs and either 60 μM dATP, 10 μM [ $\alpha$ -<sup>32</sup>P]dTTP, or 60 μM dTTP, 10 μM [ $\alpha$ -<sup>32</sup>P]dATP, and 200 nM primase or DNA 20mer (unless indicated otherwise) for 35 s. For S1 analysis, reactions were quenched, and beads washed twice in Buffer E and once in S1 buffer. Beads were re-suspended in 20 μl S1 buffer. Five-microlitre beads were incubated with 1 U S1 for 7.5 min in 15 μl. Single-molecule total internal reflection fluorescence microscopy was performed as described<sup>16</sup>, and is further elaborated in Methods.

**Full Methods** and any associated references are available in the online version of the paper.

Received 11 May 2012; accepted 7 February 2013.

Published online 27 March 2013.

1. Kornberg, A. & Baker, T. A. *DNA Replication* 2nd edn (W. H. Freeman, 1992).
2. Ullsperger, C. J., Volgodskii, A. V. & Cozzarelli, N. R. *Unlinking of DNA by Topoisomerases during DNA Replication* Vol. 9 (Springer, 1995).
3. Hingorani, M. M. & O'Donnell, M. Sliding clamps: a (tail)ored fit. *Curr. Biol.* **10**, R25–R29 (2000).

4. Wyman, C. & Botchan, M. DNA replication. A familiar ring to DNA polymerase processivity. *Curr. Biol.* **5**, 334–337 (1995).
5. Hamdan, S. M., Loparo, J. J., Takahashi, M., Richardson, C. C. & van Oijen, A. M. Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* **457**, 336–339 (2009).
6. Li, X. & Marians, K. J. Two distinct triggers for cycling of the lagging strand polymerase at the replication fork. *J. Biol. Chem.* **275**, 34757–34765 (2000).
7. Yang, J., Nelson, S. W. & Benkovic, S. J. The control mechanism for lagging strand polymerase recycling during bacteriophage T4 DNA replication. *Mol. Cell* **21**, 153–164 (2006).
8. Johnson, A. & O'Donnell, M. Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* **74**, 283–315 (2005).
9. Postow, L., Peter, B. J. & Cozzarelli, N. R. Knot what we thought before: the twisted story of replication. *Bioessays* **21**, 805–808 (1999).
10. Chastain, P. D. II, Makhov, A. M., Nossal, N. G. & Griffith, J. Architecture of the replication complex and DNA loops at the fork generated by the bacteriophage t4 proteins. *J. Biol. Chem.* **278**, 21276–21285 (2003).
11. Volgodskii, A. V., Lukashin, A. V., Anshelevich, V. V. & Frank-Kamenetskii, M. D. Fluctuations in superhelical DNA. *Nucleic Acids Res.* **6**, 967–982 (1979).
12. Leu, F. P., Georgescu, R. & O'Donnell, M. Mechanism of the *E. coli* tau processivity switch during lagging-strand synthesis. *Mol. Cell* **11**, 315–327 (2003).
13. Espeli, O. & Marians, K. J. Untangling intracellular DNA topology. *Mol. Microbiol.* **52**, 925–931 (2004).
14. Stockum, A., Lloyd, R. G. & Rudolph, C. J. On the viability of *Escherichia coli* cells lacking DNA topoisomerase I. *BMC Microbiol.* **12**, 26 (2012).
15. Stupina, V. A. & Wang, J. C. Viability of *Escherichia coli* topA mutants lacking DNA topoisomerase I. *J. Biol. Chem.* **280**, 355–360 (2005).
16. Yao, N. Y., Georgescu, R. E., Finkelstein, J. & O'Donnell, M. E. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proc. Natl Acad. Sci. USA* **106**, 13236–13241 (2009).
17. Tanner, N. A. et al. Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic Acids Res.* **37**, e27 (2009).
18. O'Donnell, M. E. & Kornberg, A. Complete replication of templates by *Escherichia coli* DNA polymerase III holoenzyme. *J. Biol. Chem.* **260**, 12884–12889 (1985).
19. McHenry, C. S. DNA replicases from a bacterial perspective. *Annu. Rev. Biochem.* **80**, 403–436 (2011).
20. Wu, C. A., Zechner, E. L., Reems, J. A., McHenry, C. S. & Marians, K. J. Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. V. Primase action regulates the cycle of Okazaki fragment synthesis. *J. Biol. Chem.* **267**, 4074–4083 (1992).
21. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
22. Dorman, C. J. DNA supercoiling and bacterial gene expression. *Sci. Prog.* **89**, 151–166 (2006).
23. Travers, A. & Muskhelishvili, G. A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep.* **8**, 147–151 (2007).
24. Benyajati, C. & Worcel, A. Isolation, characterization, and structure of the folded interphase genome of *Drosophila melanogaster*. *Cell* **9**, 393–407 (1976).
25. McInerney, P. & O'Donnell, M. Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. *J. Biol. Chem.* **279**, 21543–21551 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank L. Langston, A. Libchaber and B. Michel for suggestions on the manuscript. We are also grateful to the National Institutes of Health (GM 38839) for supporting this work.

**Author Contributions** I.K. and M.O.D. conceived the project, I.K. and R.G. performed experiments, and I.K., R.G. and M.O.D. designed the experiments, analysed data and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.O.D. ([odonnel@rockefeller.edu](mailto:odonnel@rockefeller.edu)).

## METHODS

**Reagents and proteins.** Proteins were purified as described:  $\alpha$ ,  $\varepsilon$ ,  $\gamma$ ,  $\tau$ ,  $\delta$ ,  $\delta'$ ,  $\chi$ ,  $\psi$ ,  $\theta$ ,  $\beta^{26}$ , SSB<sup>26</sup>, Pol III\* ( $(\alpha\varepsilon\theta)_3\tau_3\delta\delta'\chi\psi$ ) and  $\gamma$  complex ( $\gamma_3\delta\delta'\chi\psi$ ) were reconstituted as described<sup>27–29</sup>. Solid support bead-based assays were performed using Pol III\* containing an  $\varepsilon$  mutant (D12A and E14A) that eliminated the 3'–5' exonuclease activity to prevent removal of the 3' dideoxyterminal nucleotide of the Trap DNA. Replication buffer was 20 mM Tris-HCl (pH 7.5), 4% glycerol, 0.1 mM EDTA, 40  $\mu$ g ml<sup>-1</sup> BSA, 5 mM DTT and 10 mM MgOAc<sub>2</sub>. S1 buffer was 40 mM NaOAc (pH 4.5), 300 mM NaCl and 2 mM ZnSO<sub>4</sub>.

**Rolling circle replication reactions.** Twenty microlitres of streptavidin-coupled magnetic beads (Invitrogen) were washed and equilibrated in replication buffer. Replisomes were assembled in 20  $\mu$ l replication buffer (20 mM Tris-HCl (pH 7.5), 4% glycerol, 0.1 mM EDTA, 40  $\mu$ g ml<sup>-1</sup> BSA, 5 mM DTT, 10 mM MgOAc<sub>2</sub>) containing 100 fmol synthetic 5'-biotinylated 100-base oligonucleotide rolling circle DNA (prepared as in<sup>25</sup>), 60  $\mu$ M each dATP and dGTP, 50  $\mu$ M ATP- $\gamma$ -S, 4 pmol DnaB<sub>6</sub>, 2.5 pmol  $\beta_2$  and 0.5 pmol Pol III\*. Reactions were incubated 5 min at 37 °C before immobilization on beads for 10 min at 25 °C. Beads were washed three times in 500  $\mu$ l replication buffer containing 30  $\mu$ M dATP and dGTP, 12 pmol  $\beta_2$  and 50  $\mu$ M ATP- $\gamma$ -S before re-suspending in 10  $\mu$ l replication buffer containing 60  $\mu$ M dCTP, 60  $\mu$ M dGTP and 2.5 pmol  $\beta_2$ . Replication was initiated by addition of 10  $\mu$ l 0.5 mM ATP, 200 nM primase (unless indicated otherwise), 480 nM SSB<sub>4</sub>, 50  $\mu$ M NTPs and either: 60  $\mu$ M dATP, 10  $\mu$ M TTP, 3  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]TTP to label the lagging-strand template (that is, leading-strand product is the lagging-strand template in a rolling circle reaction), or 60  $\mu$ M dTTP, 10  $\mu$ M dATP and 3  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]dATP to label lagging-strand Okazaki fragments. After 15 s, 5  $\mu$ l replication buffer containing 20 nM Trap DNA was added to sequester any polymerase that dissociates during replication, thereby preventing it from filling gaps produced by signal release. Indicated concentrations reflect final reaction conditions in 25  $\mu$ l. Replication was allowed to proceed for 35 s. The Trap DNA consists of M13mp18 ssDNA primed with a 3' dideoxyterminated DNA 30-base oligonucleotide (5'-GTTAAAGGCCGTTTGCGGGATCGTCACddC-3') onto which the  $\beta$  clamp was pre-assembled in a 2-min reaction at 37 °C containing 4.8 pmol SSB<sub>4</sub>, 2.5 pmol  $\beta_2$ , 1 pmol  $\gamma$ -complex. Reactions were quenched by adding 25  $\mu$ l of 40 mM EDTA and 1% SDS. Where indicated, primase was substituted with 20-base DNA oligonucleotides (5'-GCAAAAGCAGGCACAGCAAC-3'). Beads were washed twice in Buffer E (20 mM Tris-HCl (pH 7.5), 4% glycerol, 0.1 mM EDTA, 40  $\mu$ g ml<sup>-1</sup> BSA, 5 mM DTT) and once in S1 buffer (200 mM sodium acetate (pH 4.5), 1.5 M NaCl and 10 mM ZnSO<sub>4</sub>).

**S1 endonuclease digestion.** Beads were re-suspended in 20  $\mu$ l S1 buffer and 5  $\mu$ l of the reaction was withdrawn and incubated with 1 U S1 nuclease (Fermentas) for 7.5 min at 37 °C in 15  $\mu$ l total reaction volume. Reactions were quenched by adding 15  $\mu$ l of 40 mM HEPES (pH 7.5), 40 mM EDTA and 6% SDS. DNA was released from beads by incubating at 95 °C for 3 min and resolved on 0.8% alkaline agarose gels.

**Control reactions for S1 assay.** Where indicated, S1 treatment was performed on either 0.2  $\mu$ g supercoiled pIK31 plasmid DNA, pIK31 linearized with PstI (pIK31-PstI), pIK31 nicked on one strand with Nb.BvCI (pIK31-BvCI) or 0.4  $\mu$ g circular M13mp18 ssDNA. Rolling circle replication reactions were performed as described above. Reactions that were treated with Pol II were first replicated and quenched as described above, washed three times in replication buffer before re-suspending in 25  $\mu$ l buffer containing 20 nM Pol II, 60  $\mu$ M dCTP and dGTP, 2.5 pmol  $\beta_2$ , 1 pmol  $\gamma$ -complex and 0.5 mM ATP and incubated for 5 min at 37 °C. Replication was initiated by adding 60  $\mu$ M dTTP, dATP and incubated for 10 min at 37 °C. Replication for fill-in reactions with Pol II to label gaps was performed similarly and initiated with dATP, dTTP and [ $\alpha$ -<sup>32</sup>P]dATP, where indicated.

**Ligation of Okazaki fragments.** Replication reactions that were treated with T4 DNA ligase (New England Biolabs) were first replicated in the presence of [ $\alpha$ -<sup>32</sup>P]dATP and 800 nM 5'-phosphorylated oligonucleotide (5'-P-GCAAAA GCAGGCACAGCAAC-3') for lagging-strand synthesis and quenched as described. Beads were washed three times in 20 mM Tris-HCl (pH 7.5), 4% glycerol, 0.1 mM EDTA, 40  $\mu$ g ml<sup>-1</sup> BSA and 5 mM DTT. Beads were re-suspended in T4 quick ligase buffer (NEB) and 1  $\mu$ l of T4 DNA quick ligase for 15 min at 25 °C. Reactions were quenched by adding 15  $\mu$ l of 40 mM HEPES (pH 7.5), 40 mM EDTA and 6% SDS. DNA was released from the beads by incubating at 95 °C for 3 min and resolved on 0.8% alkaline agarose gels.

**Analysis of lagging-strand polymerase processivity on primed  $\phi$ X174 ssDNA in trans.** Replisomes were assembled on the biotinylated rolling circle substrate as described above. After the wash steps, beads were re-suspended in 20  $\mu$ l replication buffer containing 8 nM  $\phi$ X174 ssDNA (5.4 kb) primed with a DNA oligonucleotide (5'-CAAGCAGTAGTAATTCCCTGCTTTATCAAG-3'), 2.4 pmol SSB<sub>4</sub>, 1 pmol  $\beta_2$ , 0.08 pmol  $\gamma$ -complex, 60  $\mu$ M dCTP and 60  $\mu$ M dGTP for 2 min at 37 °C. Beads were immobilized, washed and replication was initiated as described above.

**Data analysis.** Lanes in alkaline gels were scanned using a phosphorimager with ImageQuant software. The intensity of radioactivity at each pixel was normalized to the corresponding molecular mass to correct for the fact that longer products incorporate more radiolabel. Size markers were used as a reference in the gel. Percentage signal release was calculated from the size difference of S1-treated leading- and lagging-strand products (fit with a single Gaussian distribution as described in the legend to Supplementary Fig. 3) using the following calculation:

$$\text{SR}(\%) = \frac{\text{Size (lg)}}{\text{Size (ld + S1)}} \times 100$$

where SR = signal release; size (lg) = size of lagging-strand product; size (ld + S1) = size of S1 digested leading-strand product (that is, lagging-strand template).

**Single-molecule replication assays.** Replisomes were assembled onto 530 fmol (10.5 nM) 5' biotinylated 100-base oligonucleotide rolling circle DNA by first incubating with 18 pmol DnaB helicase (365 nM) in 50  $\mu$ l Buffer A (20 mM Tris-HCl, pH 7.5, 5 mM DTT, 40  $\mu$ g ml<sup>-1</sup> BSA, 4% glycerol) containing 8 mM MgOAc<sub>2</sub> and 0.25 mM ATP, followed by incubation for 30 s at 37 °C. Then a 25  $\mu$ l solution of Buffer A containing Pol III\* (675 fmol ( $\alpha\varepsilon\theta$ )<sub>3</sub> $\tau_3\delta\delta'\psi\gamma$ , 27 nM as complex),  $\beta_2$  (1.85 pmol, 74 nM as dimer), 60  $\mu$ M each dCTP and dGTP and 8 mM MgOAc<sub>2</sub> was added followed by a further 1 min at 37 °C. To immobilize the replisome-DNA complex in the flow cell, 1  $\mu$ l of the above reaction was diluted 1000-fold into 1 ml of Buffer B (8 mM MgOAc<sub>2</sub>, 60  $\mu$ M each of dCTP and dGTP, and 50 nM Yo-Pro1 in Buffer A), then the diluted reaction was passed through the flow cell at 500  $\mu$ l min<sup>-1</sup> for 30 s, and a further 2 min at 10  $\mu$ l min<sup>-1</sup>. DNA replication was initiated upon flowing Buffer A containing 60  $\mu$ M of each deoxyribonucleotide triphosphate, 250  $\mu$ M each of CTP, TTP and UTP, 1 mM ATP, 462 nM SSB<sub>4</sub>, 100 nM primase, 50 nM  $\beta_2$ , 50 mM potassium glutamate, 50 nM Yo-Pro1, 0.8% glucose, 0.01%  $\beta$ -mercaptoethanol, 0.57 U glucose oxidase and 2.1 U catalase. The flow rate was either 10  $\mu$ l min<sup>-1</sup> or 100  $\mu$ l min<sup>-1</sup>, as indicated. After 10 min of replication, the flow rate was adjusted to 100  $\mu$ l min<sup>-1</sup> to analyse replication products.

**Intensity analysis of the DNA products.** To correct for flow disturbances in our analysis, we averaged between 30 and 40 frames for each DNA strand (over two adjacent pixels); the same operation was performed for an adjacent empty region to serve as background correction that took into consideration the signal variation of non-uniform intensities shown by the bilayer. In addition to this background subtraction, we performed a signal normalization. As the DNA strand grows longer and extends farther away from the biotin-streptavidin anchoring point in the lipid bilayer, there is an additional 'slow-drift' in intensity that is due to higher strand mobility into the flow. Therefore we normalized the slow-drift intensity signal to 100% dsDNA by applying a rank-order filter, a method best suited for removing shot noise. This method finds the specified percentile of data points in the data window around each point in the data set (here the corrected line plot intensity) and replaces that point with the percentile. We analysed the data using Matlab software.

26. Yao, N., Hurwitz, J. & O'Donnell, M. Dynamics of beta and proliferating cell nuclear antigen sliding clamps in traversing DNA secondary structure. *J. Biol. Chem.* **275**, 1421–1432 (2000).
27. Onrust, R., Finkelstein, J., Turner, J., Naktinis, V. & O'Donnell, M. Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. III. Interface between two polymerases and the clamp loader. *J. Biol. Chem.* **270**, 13366–13377 (1995).
28. Bailey, S., Wing, R. A. & Steitz, T. A. The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell* **126**, 893–904 (2006).
29. Georgescu, R. E. et al. Mechanism of polymerase collision release from sliding clamps on the lagging strand. *EMBO J.* **28**, 2981–2991 (2009).

# Structures of protein–protein complexes involved in electron transfer

Svetlana V. Antonyuk<sup>1\*</sup>, Cong Han<sup>1\*</sup>, Robert R. Eady<sup>1</sup> & S. Samar Hasnain<sup>1</sup>

**Electron transfer reactions are essential for life because they underpin oxidative phosphorylation and photosynthesis, processes leading to the generation of ATP, and are involved in many reactions of intermediary metabolism<sup>1</sup>. Key to these roles is the formation of transient inter-protein electron transfer complexes. The structural basis for the control of specificity between partner proteins is lacking because these weak transient complexes have remained largely intractable for crystallographic studies<sup>2,3</sup>. Inter-protein electron transfer processes are central to all of the key steps of denitrification, an alternative form of respiration in which bacteria reduce nitrate or nitrite to N<sub>2</sub> through the gaseous intermediates nitric oxide (NO) and nitrous oxide (N<sub>2</sub>O) when oxygen concentrations are limiting. The one-electron reduction of nitrite to NO, a precursor to N<sub>2</sub>O, is performed by either a haem- or copper-containing nitrite reductase (CuNiR) where they receive an electron from redox partner proteins a cupredoxin or a c-type cytochrome<sup>4,5</sup>. Here we report the structures of the newly characterized three-domain haem-c-Cu nitrite reductase from *Ralstonia pickettii* (*Rp*NiR) at 1.01 Å resolution and its M92A and P93A mutants. Very high resolution provides the first view of the atomic detail of the interface between the core trimeric cupredoxin structure of CuNiR and the tethered cytochrome c domain that allows the enzyme to function as an effective self-electron transfer system where the donor and acceptor proteins are fused together by genomic acquisition for functional advantage. Comparison of *Rp*NiR with the binary complex of a CuNiR with a donor protein, AxNiR-cytC<sub>551</sub> (ref. 6), and mutagenesis studies provide direct evidence for the importance of a hydrogen-bonded water at the interface in electron transfer. The structure also provides an explanation for the preferential binding of nitrite to the reduced copper ion at the active site in *Rp*NiR, in contrast to other CuNiRs where reductive inactivation occurs, preventing substrate binding.**

Conversion of nitrate to N<sub>2</sub>, known as the denitrification process, is performed by several distinct enzymes<sup>4,5</sup>. The first committed step, NO<sub>2</sub><sup>-</sup> + 2H<sup>+</sup> + e<sup>-</sup> → NO + H<sub>2</sub>O is catalysed by NiR. The well-studied two-domain CuNiRs are trimers of 106 kDa, with each monomer having two domains with a characteristic β-sandwich cupredoxin motif. Each monomer has a type-1 Cu (T1Cu) centre with (Cys-Met-His<sub>2</sub>) ligation and a type-2 Cu (T2Cu) site with (His<sub>3</sub>-H<sub>2</sub>O) ligation at the interface between two monomers, both of which provide its ligands. During catalysis nitrite binds to the T2Cu, displacing the liganded H<sub>2</sub>O. Electrons produced by respiration are transferred to the T1Cu site by a partner redox protein. Subsequently electrons are transferred by a 12.6 Å internal His-Cys bridge in a proton-gated reaction, to the T2Cu site, where nitrite is reduced to NO<sup>7-12</sup>. The timing of electron and substrate delivery to the active site of CuNiRs is critical to catalysis, because if the water coordinated at the T2Cu site of the oxidized enzyme dissociates before nitrite binds to displace this water, the enzyme is inactivated<sup>13,14</sup>.

Recently, two new sub-classes of CuNiRs have been recognized that retain the core structure of two-domain enzymes but contain an extra

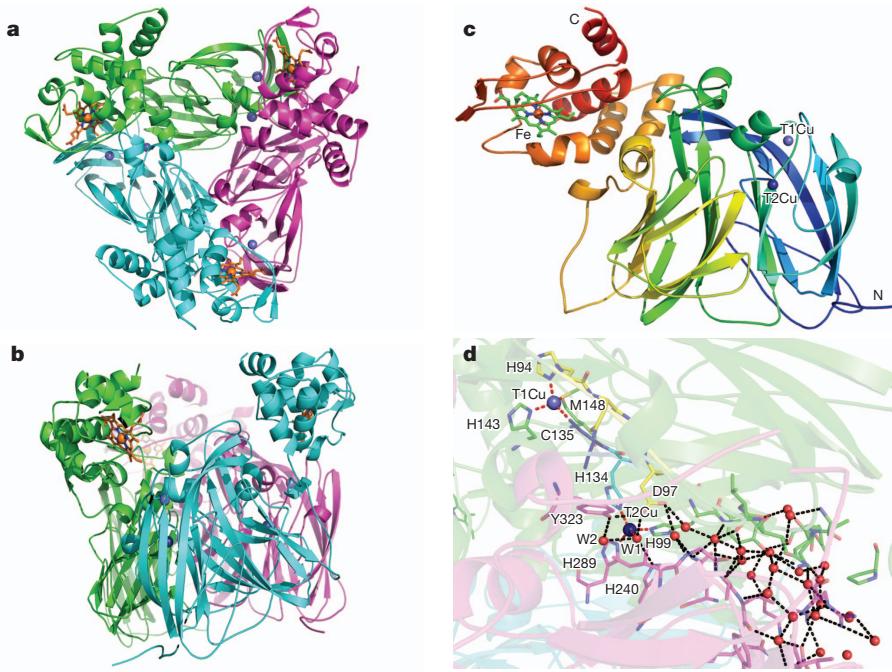
cupredoxin or haem-c-containing domain, fused at the amino (N)- and carboxy (C)-terminals respectively<sup>15,16</sup>. Genome analysis has shown that these variant enzymes are quite widely distributed in Nature. The cupredoxin N-terminal-fused three-domain NiR from *Hypomicrobium denitrificans* (*Hd*NiR) has been biochemically and structurally defined<sup>17,18</sup>. Surprisingly the observed hexameric structure, against all expectations, showed that the Cu site in the extra cupredoxin domain was located too far away from the catalytic core for effective electron transfer<sup>17</sup>.

The first biochemical and functional characterization of a fused three-domain cytochrome c CuNiR, the enzyme from *R. pickettii* (*Rp*NiR)<sup>19</sup>, showed it to be trimeric, similar to two-domain CuNiRs such as AxNiR. However, *Rp*NiR showed a major difference in that the oxidized resting state had a significantly lower nitrite binding affinity. Our structure of *Rp*NiR at 1.01 Å resolution reported here (Fig. 1) provides the structural basis of the functional importance of the haem c in *Rp*NiR, where electron transfer from the reduced haem occurs during turnover. It also provides an explanation as to how the coordinated water at the T2Cu site is stabilized, preventing its displacement by nitrite, rationalizing the low binding affinity of nitrite to the oxidized enzyme. A comparison of the *Rp*NiR structure with the cupredoxin N-terminal fused NiR, *Hd*NiR, shows a remarkable conservation of the three-dimensional structural organization for these two new classes of extended three-domain NiRs (Supplementary Fig. 1). T1Cu of the cupredoxin domain of *Hd*NiR can be brought to an appropriate distance for electron transfer to the catalytic T1Cu-T2Cu core by a simple rotation of the cupredoxin domain.

Potential electron transfer routes between cupredoxin and haem c domains have been clarified to some extent by structural studies of several other electron transfer complexes: the transient binary AxNiR-CytC<sub>551</sub> complex (1.7 Å resolution)<sup>6</sup> and the integral membrane complex caa<sub>3</sub>-type cytochrome c oxidase (2.36 Å resolution)<sup>20</sup>, for example. Our structure of *Rp*NiR clearly defines electron transfer routes between a haem c domain and a catalytic cupredoxin core. The fold of the trimeric CuNiR core domain is preserved and the haem c domain of one monomer is in close proximity to the T1Cu site of another monomer at an electron transfer compatible distance of 10.1 Å between the T1Cu ion and the closest haem edge (CBC methyl carbon) of the porphyrin (Fig. 2). The linker between haem c and core NiR domain forms ten salt bridges, eleven hydrogen bonds with the core and nine hydrogen bonds with the haem domain contributing to the stability of the haem c domain. These interactions would preclude any significant movement of the haem domain from its position that favours electron transfer to the T1Cu centre. Comparison of *Rp*NiR with the structure of the related but biochemically uncharacterized enzyme from *Pseudoalteromonas haloplanktis* (*Ph*NiR) (Protein Data Bank entry 2zoo) showed differences in the overall organization of the subunits. The *Rp*NiR monomer is more compact, the linker between haem and Cu binding domains (residues 315–333) is arranged in a helical structure and only two monomers are involved in forming the T2Cu site and the interface between the haem and cupredoxin

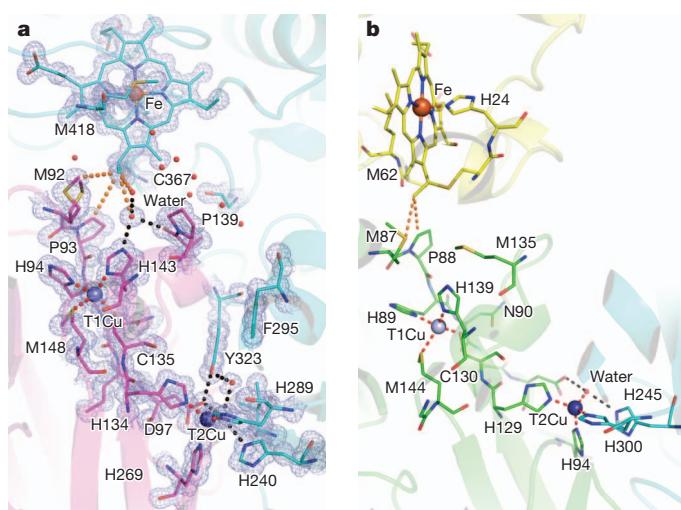
<sup>1</sup>Molecular Biophysics Group, Institute of Integrative Biology, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 7ZK, UK.

\*These authors contributed equally to this work.



**Figure 1 | Structural organization of RpNiR.** **a, b**, Trimer viewed along (a) and perpendicular to (b) the three-fold axis. T2Cu (darker blue sphere) binds at the interface between two monomers (a, b). T1Cu (lighter blue sphere) is located closer to the surface of each monomer in the proximity of the cytochrome domain from the adjacent monomer at a distance commensurate with efficient electron transfer. **c**, Ribbon diagram of a monomer with

domains. In contrast, in *PhNiR* the linker wraps around the neighbouring monomer and reaches the distant third monomer, creating the cytochrome and catalytic domain interface (Supplementary Fig. 1). The significance of these differences awaits the biochemical characterization of *PhNiR*.

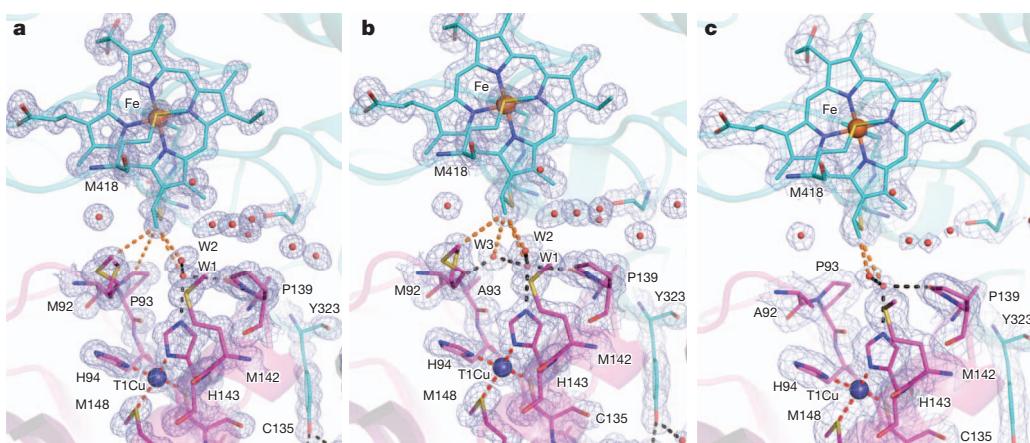


**Figure 2 | Details of interaction between cytochrome and Cu binding domains.** **a**, RpNiR. An electron transfer route can be traced from the porphyrin CBC atom by a water molecule, the Ne2 atom of the His 143 ligand of the T1Cu centre. For the RpNiR, a  $2F_o - F_c$  map contoured at  $1.0\sigma$  for the T1Cu and T2Cu sites is shown. Fe-T1Cu distance is  $10.4\text{ \AA}$ , T1-T2Cu is  $12.7\text{ \AA}$ . For simplicity, only one conserved water molecule on the shortest route to T1Cu is shown at the interface between the cytochrome domain and T1Cu. This water is located in very close proximity both to the CBC haem atom ( $3.6\text{ \AA}$ ) and the Ne2 atom of the Cu-ligating His 143 ( $2.8\text{ \AA}$ ). **b**, AxNIR-Cytc<sub>551</sub> complex<sup>6</sup> with shortest distances between haem and Cu binding domain shown as dotted orange lines. The binary complex, AxNIR-Cytc<sub>551</sub>, has a hydrophobic interface between cytochrome c<sub>551</sub> and the T1Cu site.

cytochrome domain (orange-red), cupredoxin domains (green-blue) containing T1Cu and T2Cu and the helical linker structure formed by residues 315–333 (yellow) between the two (c) and details of the water (red) channel linking it to W1 coordinating to T2Cu (d). The stability of the haem c domain in RpNiR is also reinforced by the extended interface area of  $2267\text{ \AA}^2$  between the two monomers in the trimer, of which  $452\text{ \AA}^2$  involves the haem c domain.

A striking feature of the RpNiR structure is the extensive water network at the interface of the haem c domain and the surface above the T1Cu of the core of the enzyme (Figs 2 and 3a). Several computational studies have identified the importance of water molecules hydrogen bonded between the donor and acceptor sites in providing superior electronic coupling. This results in substantially enhanced rates of electron transfer, as between methylamine dehydrogenase and the cupredoxin amicyanin<sup>21</sup>. In the case of RpNiR, the interactions between the cytochrome and the catalytic-core domains are either C–C interactions of haem with Met 92 and Pro 93 or are mediated by water hydrogen bonds between the two domains. Similar interactions are seen in *PhNiR* (Supplementary Fig. 2). In contrast to these hydrated interfaces, the structure of the transient binary AxNiR-Cytc<sub>551</sub> complex shows that the protein–protein interface is a primarily hydrophobic patch with a non-polar core sealed off from the aqueous environment<sup>6</sup>, with no water molecules that may be involved in electron transfer. In this complex, and in native AxNiR, His 135 (His 143 in RpNiR) is protected from hydrophilic environment by Met 135 and Trp 138. A similar situation is found in the tethered membrane complex *caa3*-type cytochrome c oxidase of *Thermus thermophilus*, where the electron transfer pathway from haem c to the Cu<sub>A</sub> centre does not involve water, but rather the D-pyrrole and D-propionate of haem c to the *cis* amide N and  $\alpha$  C of Phe 126 of the cupredoxin domain<sup>20</sup>.

The potential involvement of Met 92 and Pro 93 in electron transfer of RpNiR was tested by individual substitution by Ala. Structures of these variants determined at  $1.9$  and  $1.4\text{ \AA}$  respectively show that the conserved water was not perturbed by these substitutions (Fig. 3). The mutations had no significant effect on the specific activity of the enzyme; and in single turnover experiments where reduced RpNiR was reoxidized by nitrite, the electron transfer efficiency from haem to T2Cu was similar to the wild-type enzyme, with M92A variant showing a small decrease in rate (see Supplementary Information). These small effects suggest that the conserved water molecule hydrogen bonded to the solvent-exposed T1Cu histidine ligand and that the carbonyl of the peptide bond of Ala 138 plays a dominant role in



**Figure 3 | Comparison of interactions between cytochrome and Cu binding domains in wild type RpNiR, P93A and M92A mutants.** **a–c,** RpNiR (a), P93A (b) and M92A (c) in H3 crystal form. A  $2F_o - F_c$  map contoured at  $1.0\sigma$  at  $1.01/1.4/1.8 \text{ \AA}$  resolution for wild type/P93A/M92A. For all three structures, the distances between Fe and T1 Cu is  $10.1\text{--}10.3 \text{ \AA}$ . The interface between the cytochrome domain and the T1Cu site has two conserved water

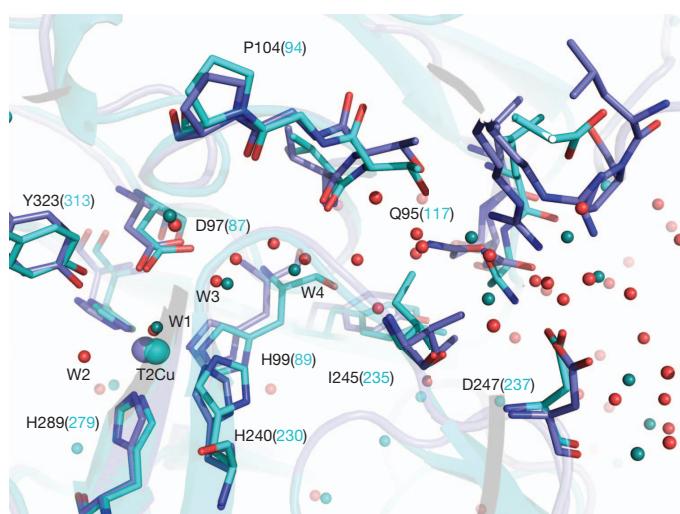
molecules (W1 and W2) located in very close proximity to the CBC haem atom ( $3.5\text{--}3.6$  and  $3.3\text{--}3.4 \text{ \AA}$  respectively); the waters are hydrogen bonded to each other. The distances from the W1 molecule to His 143(Ne2) and Ala 138(O) are approximately  $2.8 \text{ \AA}$  in three structures. Extra water W3 is seen in P93A close to the mutation site.

electron transfer in this tethered complex. The nature of this interaction precludes further testing of its involvement in electron transfer by mutation. This role for hydrogen-bonded water in electron transfer between the haem and T1Cu site contrasts with the transient binary *AxNiR-Cytc<sub>551</sub>* complex, where the close contact between the two proteins results in use of C–C interactions between the CBC methyl group and Pro 88 of the cupredoxin domain of the core NiR. These two electron transfer systems illustrate the two classes of predicted electron tunnelling processes: protein mediated and structured water mediated<sup>22</sup>.

The structure also provides insights into the mechanism of this recently characterized class of CuNiRs. A surprising feature of the RpNiR structure was that, despite the similarity to two domain NiRs, the usual substrate access channel is closed. Atomic-resolution structures of two-domain NiRs have shown the detail of two proton pathways to the active site and the hydrophobic substrate access channel to

the protein surface<sup>8,9</sup>. Mutagenesis studies of *AxNiR* established only one proton pathway to be functionally relevant, namely the highly ordered (Asp 92–water–water–Ala 131–Asn 90–Asn 107) network from the T2Cu site (Supplementary Fig. 3) to the protein surface<sup>23</sup>. This hydrophobic channel, approximately  $6 \text{ \AA}$  wide formed by residues belonging to adjacent monomers, was also assigned as the route of substrate access to the catalytic centre. In RpNiR this channel extends towards the haem domain, but lacks water molecules as it is blocked by residue Tyr 323 that forms part of the linker between cytochrome and cupredoxin domains. Instead, the non-functional proton pathway of two domain NiRs predominates (Fig. 4 and Supplementary Fig. 1d). This channel is present in the structure of *PhNiR* also. However, in *PhNiR*, Ile 235 adopts a different orientation to Ile 245 of RpNiR and interrupts it. This residue may thus be a channel activator and the two structures may represent two alternative protein states.

In contrast to other CuNiRs, the oxidized T2Cu catalytic sites of RpNiR have a very low affinity for nitrite<sup>19</sup>. This does not arise from constraints imposed by the altered substrate access channel described above because the apparent Michaelis constant for nitrite is  $26 \mu\text{M}$ , comparable to other CuNiRs. Structural and kinetic studies of CuNiRs indicate that reduction of the T2Cu can result in the loss of the coordinated  $\text{H}_2\text{O}$  ligand to form a catalytically inactive  $\text{Cu}^+(\text{His})_3$  coordination at the active site<sup>13,14</sup>. The water ligated to the T2Cu of NiR in all other structures hydrogen bonds to the carboxylate group of the active pocket Asp at  $2.54\text{--}2.83 \text{ \AA}$  and a histidyl nitrogen atom<sup>8,9,24</sup>. In RpNiR the coordinated water W1 has two extra hydrogen bonds (Supplementary Fig. 4), being connected to a second water molecule (labelled W2) as well as Oδ1 of Asp 97 that is also hydrogen bonded to the OH of Tyr 323. Water W2 also has a hydrogen bond to the OH of Tyr 323 that blocks the putative substrate entry channel identified in two-domain NiRs. This would be expected to provide more stability to the coordinated water molecule, hindering nitrite binding to the oxidized site but stabilizing it on reduction to allow nitrite to bind by displacement of the water ligand before the catalytically inactive  $\text{Cu}^+(\text{His})_3$  coordination species is formed.



**Figure 4 | Details of the proton pathway in RpNiR and its comparison with PhNiR (Protein Data Bank entry 2zoo).** RpNiR is shown in dark blue with water as red sphere, PhNiR in light blue. This channel comprises residues on the interface between two monomers that extends from the T2Cu catalytic centre to the protein surface and is filled with an extensive water network. In the case of PhNiR, the channel is interrupted by Ile 235, which has a different orientation to the corresponding Ile 245 of RpNiR. Ile 235/245 in these extended NiRs correspond to the conserved His 254 residue in the two-domain *AxNiR*<sup>23</sup>.

## METHODS SUMMARY

*RpNiR* was expressed in *Escherichia coli* BL21(DE3) and purified as previously described<sup>19</sup>. For M92A and P92A, site-directed mutagenesis was performed following the QuikChange site-directed mutagenesis protocol (Agilent), using KOD Hot Start DNA polymerase (Merck). The plasmid pET26b-RpNiR was used as template for introducing the following single amino-acid changes in RpNiR by site-directed mutagenesis: M92A and P93A. The primers for the M92A mutant were (sense) 5'-CACCCGAGCAGCAAGGC CGCGACAACATCGA-3' and

(antisense) 5'-TCGATGTTGTGGCGCCCTTGCTGCTCGGGTG-3'. The primers for the P93A mutant were (sense) 5'-CGAGCAGCAAGATGGCGCAC AACATCGAC-3' and (antisense) 5'-GTCGATGTTGTGCGCCATCTGCTG CTCG-3'. The mutant plasmids were confirmed by DNA sequencing before transformation into *E. coli* strain BL21(DE3). Both mutants were purified following native protein protocol. Crystals for RpNiR and its mutants (H3 crystal form) were grown over a month using the hanging drop vapour diffusion method at 4 °C. Wild-type RpNiR P2<sub>1</sub>3 crystal form grew in 4 days. Diffraction data were collected from the single crystal at 100 K using a PILATUS-6M detector at Proxima1 (Soleil), at ID24 and IO2 (Diamond). For native protein crystals (RpNiR-H3/RpNiR-P2<sub>1</sub>3), four data sets were collected at 0.98, 1.22, 1.33 and 1.7 Å X-ray wavelengths to confirm correct metal incorporation. All data were processed and scaled by HKL2000 software<sup>25</sup>. Only data sets collected at the 0.98 Å wavelength were used for structural determination and refinement. Native RpNiR-H3 structure was solved by molecular replacement using MOLREP<sup>26</sup> in the CCP4 (ref. 27) program suite and PhNiR (Protein Data Bank entry 2zoo) as the search model. H3 and P2<sub>1</sub>3 forms have one and four monomers, respectively, in an asymmetric unit. RpNiR-P2<sub>1</sub>3 was solved by molecular replacement using an RpNiR-H3 model. All structures were refined using REFMAC5 (ref. 28) in the CCP4 program suite. Rebuilding of the model between refinement cycles and adding waters was performed in Coot<sup>29</sup>. Data collection and refinement statistics are summarized in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper.

Received 10 June 2012; accepted 8 February 2013.

Published online 27 March 2013.

- Moser, C. M., Keske, M., Warnke, K., Farid, R. S. & Dutton, P. L. Nature of biological electron transfer. *Nature* **355**, 796–802 (1992).
- Jeng, M.-F., Englander, S. W., Pardue, K., Rogalskyj, J. S. & McLendon, G. Structural dynamics in an electron-transfer complex. *Nature Struct. Biol.* **1**, 234–238 (1994).
- Williams, P. A. *et al.* Pseudospecific docking surfaces on electron transfer proteins as illustrated by pseudoazurin, cytochrome c<sub>550</sub> and cytochrome cd<sub>1</sub> nitrite reductase. *Nature Struct. Biol.* **2**, 975–982 (1995).
- Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
- Eady, R. R. & Hasnain, S. S. In *Comprehensive Coordination Chemistry II* Vol. 8 (eds Que, L. & Tolman, W.), Ch. 28 759–786 (Elsevier, 2003).
- Nojiri, M. *et al.* Structural basis of inter-protein electron transfer for nitrite reduction in denitrification. *Nature* **462**, 117–121 (2009).
- Merkle, A. C. & Lehnert, N. Binding and activation of nitrite and nitric oxide by copper nitrite reductase and corresponding model complexes. *Dalton Trans.* **41**, 3355–3368 (2012).
- Ellis, M. J., Dodd, F. E., Sawers, G., Eady, R. R. & Hasnain, S. S. Atomic resolution structures of native copper nitrite reductase from *Alcaligenes xylosoxidans* and the active site mutant Asp92Glu. *J. Mol. Biol.* **328**, 429–438 (2003).
- Antonyuk, S. V., Strange, R. W., Sawers, G., Eady, R. R. & Hasnain, S. S. Atomic resolution structures of resting-state, substrate- and product-complexed Cu-nitrite reductase provide insight into catalytic mechanism. *Proc. Natl Acad. Sci. USA* **102**, 12041–12046 (2005).
- Godden, J. W. *et al.* The 2.3 angstrom Xray structure of nitrite reductase from *Achromobacter cycloclastes*. *Science* **253**, 438–442 (1991).
- Tocheva, E. I., Rosell, F. I., Mauk, A. G. & Murphy, M. E. P. Side-on copper-nitrosyl coordination by nitrite reductase. *Science* **304**, 867–870 (2004).
- Strange, R. W. *et al.* The substrate binding site in Cu nitrite reductase and its similarity to Zn carbonic anhydrase. *Nature Struct. Biol.* **2**, 287–292 (1995).
- Strange, R. W. *et al.* Structural and kinetic evidence for an ordered mechanism of copper nitrite reductase. *J. Mol. Biol.* **287**, 1001–1009 (1999).
- Wijma, H. J., Jeuken, L. J. C., Verbeet, M., Ph., Armstrong, F. A. & Canters, G. W. Protein film voltammetry of copper-containing nitrite reductase reveals reversible inactivation. *J. Am. Chem. Soc.* **128**, 8557–8565 (2007).
- Bertini, I. I. & Cavallaro, G. G. Cytochrome c: occurrence and functions. *Chem. Rev.* **106**, 90–115 (2006).
- Ellis, M. J., Grossmann, J. G., Eady, R. R. & Hasnain, S. S. Genomic analysis reveals widespread occurrence of new classes of copper nitrite reductases. *J. Biol. Inorg. Chem.* **12**, 1119–1127 (2007).
- Nojiri, M. *et al.* Structure and function of a hexameric copper-containing nitrite reductase. *Proc. Natl Acad. Sci. USA* **104**, 4315–4320 (2007).
- Yamaguchi, K. *et al.* Characterization of two type 1 Cu sites of *Hyphomicrobium denitrificans* nitrite reductase: a new class of copper-containing nitrite reductases. *Biochemistry* **43**, 14180–14188 (2004).
- Han, C. *et al.* Characterization of a novel copper heme c dissimilatory nitrite reductase from *Ralstonia pickettii*. *Biochem. J.* **444**, 219–226 (2012).
- Lyons, J. A. *et al.* Structural insights into electron transfer in caa<sub>3</sub>-type cytochrome oxidase. *Nature* **487**, 514–518 (2012).
- de la Lande, A., Babcock, N. S., Rezac, J., Sanders, B. C. & Salahub, D. R. Surface residues dynamically organize water bridges to enhance electron transfer between proteins. *Proc. Natl Acad. Sci. USA* **107**, 11799–11804 (2010).
- Lin, J., Balabin, A. & Beratan, D. N. The nature of aqueous tunneling pathways between electron-transfer proteins. *Science* **310**, 1311–1313 (2005).
- Hough, M. A., Eady, R. R. & Hasnain, S. S. Identification of the proton channel to the active site type 2 Cu centre of nitrite reductase: structural and enzymatic properties of His254Phe and Asn90Ser mutants. *Biochemistry* **47**, 13547–13553 (2008).
- Boulanger, M. J., Kukimoto, M., Nishiyama, M., Horinouchi, S. & Murphy, M. E. Catalytic roles for two water bridged residues (Asp-98 and His-255) in the active site of copper-containing nitrite reductase. *J. Biol. Chem.* **275**, 23957–23964 (2000).
- Otwinowski, Z. & Minor, W. In *Methods in Enzymology: Macromolecular Crystallography A* (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic Press, 1997).
- Vagin, A. A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
- Collaborative Computational Project, Number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the Biotechnology and Biological Sciences Research Council, UK (grant number BB/G005869/1 (to S.S.H. & R.R.E.)). S.V.A. acknowledges the support from the Wellcome Trust (grant number 097826/Z/11/Z). We thank the staff and management of SOLEIL and Diamond for the provision of crystallographic facilities at their synchrotron centres. We thank the members of the molecular biophysics groups, particularly R. Strange and G. Wright, for their help.

**Author Contributions** S.V.A., R.R.E. and S.S.H. conceived and designed the project; C.H. cloned, expressed and purified proteins; S.V.A. and C.H. crystallized the proteins; S.V.A. did data processing, structure determination and refinement; S.V.A., R.R.E. and S.S.H. wrote the manuscript.

**Author Information** Atomic coordinates and structure factors for the crystal structures have been deposited in Protein Data Bank under accession numbers 3ziy (r3ziysf), 4ax3 (r4ax3sf), 2yqb (r2yqbsf) and 3zbm (r3zbmsf). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S.H. (s.s.hasnain@liverpool.ac.uk) or S.V.A. (antonyuk@liverpool.ac.uk).

## METHODS

*RpNiR* was purified by DEAE ion-exchange chromatography and size-exclusion chromatography. The size-exclusion chromatography buffer was 20 mM Tris-HCl, pH 8.0, 200 mM NaCl. The *RpNiR* protein was dialysed against size-exclusion chromatography buffer plus 0.1 mM CuSO<sub>4</sub> for reconstitution of type 2 copper sites. The fraction was then dialysed exhaustively against three changes of 20 mM Tris-HCl, pH 7.5, 200 mM NaCl with a minimum equilibration period of 5 h between buffer changes. The protein samples were pooled and concentrated up to 15 mg ml<sup>-1</sup>.

Crystals for *RpNiR* and its mutants (H3 crystal form) were grown using the hanging-drop vapour-diffusion method at 4 °C temperature from similar crystallization conditions: 2 ml of protein solution in 20 mM Tris-HCl, 200 mM NaCl (pH 8.0) was mixed with an equal volume of reservoir solution containing 20% PEG3350, 0.2 M sodium citrate. The buffer for reservoir solution was 20 mM Tris-HCl, pH 7.5 for wild-type protein and 20 mM MES pH 6.5 for the mutants. Protein concentration was approximately 15 mg ml<sup>-1</sup> H3 crystal form. Wild-type *RpNiR* P2<sub>1</sub>3 crystal form was grown from 20% PEG3350, 0.2 sodium citrate, 200 mM NaCl and protein concentration 7.5 mg ml<sup>-1</sup>. All crystals were flashed cooled in the mother liquor plus 10% glycerol solution.

For *RpNiR*-H3 and P93A-H3 structures, anisotropic temperature factors were refined and riding hydrogen atoms were added to the model. The lower-resolution M93A-H3 structure was refined with isotropic B-factors and riding hydrogen atoms in the calculated positions. No restraints were used for Cu atoms refinement. The quality of the models was assessed using Coot<sup>29</sup>, Procheck<sup>30</sup> and the Molprobity<sup>31</sup> server. Comparison of the C $\alpha$  atom positions in all *RpNiR* structures (*RpNiR*-H3, *RpNiR*-P2<sub>1</sub>3) showed average root mean squared deviation between the C $\alpha$  atoms of approximately 0.2 Å, with higher flexibility in the N terminus (residues 1–16) and part of cytochrome domain (residues 383–393) atoms. The first three residues of the N terminus and the last six residues of the C terminus were not visible in the electron density map. Within five monomers in two structures (*RpNiR*-H3, *RpNiR*-P2<sub>1</sub>3), the position of the cytochrome domain was identical and crystal contacts independent. Analyses of anomalous signal in the Cu and Fe sites confirmed correct Cu and Fe incorporation.

30. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. App. Crystallogr.* **26**, 283–291 (1993).
31. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

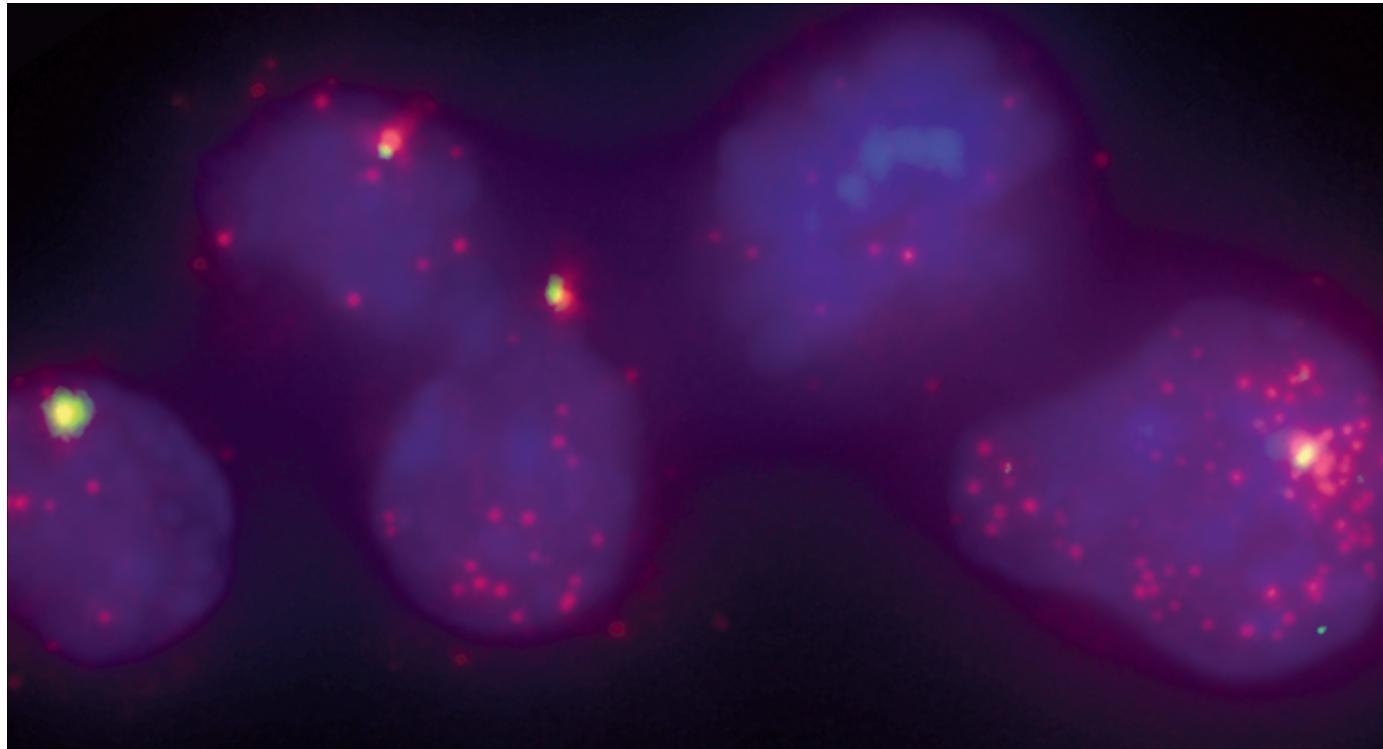
# CAREERS

**COLUMN** A lab retreat offers a chance to reinvigorate a research programme p.129

**CAREERS BLOG** The latest discussions and news on research jobs [go.nature.com/z8g4a7](http://go.nature.com/z8g4a7)

 **NATURE JOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

JOHN RINN/ARJUN RAJ



Long non-coding RNA (red) regulates DNA and proteins in the nuclei of cells.

RNA

## The genome's rising stars

*As the once-fringe field of long non-coding RNA moves into the limelight, young researchers could reap the benefits.*

BY AMY MAXMEN

When Saba Valadkhan lingered in the hallways at conferences, absorbed in discussions about the strands of 'junk' DNA that litter the human genome, she was not looking for work. She was consumed with curiosity about the possibility that long RNA sequences that do not encode proteins nevertheless have a function — enhancing or suppressing gene expression. Valadkhan's enthusiasm about the budding field of long non-coding RNA (lncRNA) did not go unnoticed: senior investigators were on the hunt for young researchers willing to pursue the topic. "Before I was even looking for job opportunities, I was told about people who were hiring,"

says Valadkhan. Soon after receiving her PhD for studies on small nuclear RNA — a type of non-coding RNA — at Columbia University in New York in 2003, she took a position as an assistant professor at Case Western Reserve University in Cleveland, Ohio.

John Rinn, now a molecular biologist at Harvard University and the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, also had a rapid career launch; as a post-doc at Stanford University in California, he was noticed at meetings where he spoke about his research on how lncRNA silences genes involved in embryonic development. Rinn was offered several faculty positions, but was sold on the Cambridge post when Stuart Schreiber, a chemical biologist and a founding member

of the Broad Institute, told him: "Every day I come to work dreaming of how I will bend the genome to my will." Rinn wanted to bend the genome with lncRNA to learn how to prevent and cure diseases.

At the time, modifying gene expression using lncRNA was not a common goal. In the early 2000s, most molecular biologists were interested in the 1–2% of the human genome that encodes proteins, which were presumed to be the brokers of biological functions. But with the rise of high-throughput sequencing, researchers learned that far from being without function, much of the rest of the genome was transcribed into non-coding RNA, including 20-nucleotide microRNAs, which suppress genes, and lncRNAs of 100 nucleotides or more. ▶

► Last September, the multi-institution ENCODE project to catalogue human DNA elements (see *Nature* **489**, 46–48; 2012) revealed that three-quarters of the human genome is transcribed into non-coding RNA, and that there may be between 10,000 and 200,000 lncRNAs. Scientists have shown that these can activate gene expression and silence genes, and links with disease have begun to emerge.

Enthusiasm for lncRNA has replaced much of the science community's scepticism. Molecular biology and biochemistry departments have taken note of a flurry of high-impact manuscripts, and some are hiring scientists to work in the emerging field. Funding is becoming easier to find. And the first biotechnology companies focused on lncRNA have taken root.

Scientists hoping to work in the field will benefit from experience with model organisms and cell cultures, or from a background in bioinformatics. But they also need traits such as creativity, a willingness to collaborate and the confidence to weather disappointment — in emerging fields, no experiments are straightforward and foolproof. For some researchers, however, these risks are part of the allure. "We are part of a very fast-moving field," says Valadkhan. "You start projects and you have no idea what to expect. It's that kind of challenge that I enjoy very much as a scientist."

Despite the wave of enthusiasm, lncRNA studies may still be difficult to launch. "RNA has really been given the shaft," says Tom Cech, a chemist at the University of Colorado Boulder who shared the 1989 Nobel Prize in Chemistry for discovering that RNA can catalyse reactions. "People were reluctant to hear

that RNA could do something other than lead to proteins back in the 1970s. But they are still surprised to hear it now."

There are no optimized experimental protocols and few clues to the habits of individual lncRNAs, so experiments often fail. And when they do work, investigators need to go the extra mile to convince reviewers that their results are real. "We had a difficult time publishing our studies for the first couple of years I was at Case Western," says Valadkhan. "It's so novel that, understandably, people are very sceptical."

"I have had so many conversations where people think I'm just full of crap," says Kevin Morris, a molecular biologist at the Scripps Research Institute in San Diego, California, who studies lncRNAs involved in HIV and cancer. "You need a thick skin to be in this field. You need to do it because you love it."

#### LOOKING UP

Nevertheless, more grants are being funded for lncRNA projects. At the US National Institutes of Health (NIH) in Bethesda, Maryland, 28 applications on the subject have been approved since the start of 2012. By contrast, just six were funded between 2008 and 2011.

Specific calls for lncRNA studies are also becoming more common. A 2012 funding announcement from the NIH's National Institute on Drug Abuse lists studies on interactions between lncRNAs and chromatin — the cluster of DNA and proteins in a cell's nucleus — as a high scientific priority. A programme announcement from the National Cancer Institute (NCI) calls for research on lncRNAs as biomarkers for cancer. Investigators can also submit projects that use lncRNA as a tool to approach other questions. For example, Kevin Howcroft, a programme director at the NCI, notes that people studying prostate cancer who might once have looked at proteins in a

signalling pathway might now propose to analyse lncRNA.

There are other funding niches. Valadkhan's work on neuronal differentiation is funded by the ALS Therapy Alliance, a consortium based in Needham, Massachusetts, that raises money for research related to the neurodegenerative disease amyotrophic lateral sclerosis. Increasing the amount of a single lncRNA in a fibroblast, a common cell in connective tissue, can transform it into a neuron, so research on lncRNA could pave the way for treatments to reverse neuronal degeneration, says Valadkhan. The California Institute for Regenerative Medicine in San Francisco, which funds stem-cell research, has also awarded grants for research into how lncRNAs can change the fate of cells.

#### COMMERCIAL INVESTMENT

The private sector has also begun to notice lncRNA. In 2011, start-up company RaNA Therapeutics in Cambridge, Massachusetts, received \$20.7 million from investors including the agriculture biotechnology firm Monsanto in St Louis, Missouri.

RaNA's goal is to manipulate lncRNA to control genes such as those that suppress tumours or cause cancer. Individual lncRNAs have specific targets, so treatments based on them might well have more predictable effects than those based on other RNAs. And they offer a rare chance to activate, rather than just block, gene expression.

In 2012, the company grew from 2 employees to 26. President and co-founder Art Krieg says that he plans to expand again in 2014 — and that he will put a premium on hiring people adept at teamwork. At the moment, RaNA's staff scientists are using high-throughput sequencing and a technique called chromatin immunoprecipitation to learn how lncRNA binds to proteins.

At least two other biotechnology companies focus on lncRNA. OPKO-CURNA, a subsidiary of OPKO Health in Miami, Florida, is investigating how to activate gene expression, for eventual use in treatments. And TransSINE Technologies in Yokohama, Japan, aims to produce lncRNA-based technologies to express antibodies and other proteins for laboratory assays.

Career success for scientists in the field requires knowledge of genomics and molecular biology. Some experiments also call for experience with model systems and cell culture, but Valadkhan, who has a background in biochemistry, learned many techniques on the fly.

Computational biology is important because it allows researchers to grapple with large data sets. Institutions including the Wellcome Trust Sanger Institute in Hinxton, UK, and the Broad Institute offer classes and workshops on bioinformatics; these typically range in length from days to weeks. Some workshops require familiarity with operating systems such as Unix and Linux and computing languages such as R, but students can learn to use such tools through books and online tutorials.



John Rinn wants to use non-coding RNA to learn how diseases can be prevented.

Universities with financially well-endowed genome centres provide easy access to sequencing and computer-savvy scientists and technicians. The leading US institutions for lncRNA research include Harvard, the University of Wisconsin-Madison, Stanford, the University of Colorado Boulder and Yale University in New Haven, Connecticut. In Europe, institutions including the University of Vienna and the nearby Research Center for Molecular Medicine of the Austrian Academy of Sciences have histories of RNA research, including lncRNA.

The Center for Life Science Technologies at the RIKEN Yokohama Institute, which opened this month, will focus in part on lncRNA. "We can develop technology for several years without the pressure of writing grants and publications because we have an institutional budget," says Piero Carninci, the first director of the centre's genomic-technology division, which will employ roughly 100 scientists, including 9 principal investigators.

Researchers investigating the role of lncRNAs in disease often collaborate with clinicians. Just such a partnership helped Claes Wahlestedt, now director of the Center for Therapeutic Innovation at the University of Miami Miller School of Medicine in Florida, to discover that a non-coding RNA drives the expression of an enzyme involved in the progression of Alzheimer's disease (M. A. Faghili *et al.* *Nature Med.* **14**, 723–730; 2008).

When choosing a lab, researchers should remember that it can be an asset to have colleagues with a range of backgrounds, says Florian Karreth, a postdoc at Harvard Medical School in Boston, Massachusetts. "You don't want 20 people with a background in microRNA." In his group, he says, "there are people with experience in apoptosis, leukaemia and DNA repair, and it's great to learn from all of them". The 20 postdocs and a handful of graduate students and technicians often confer when starting experiments, and help each other to learn.

In the absence of a rich body of literature, ideas are often exchanged at conferences on RNA, epigenetics and genomics. As Valadkhan discovered, these are also good places to find jobs: senior scientists who attend may be looking for young investigators with creative ideas. Human lncRNAs are yet to be catalogued, and everyone wants to know more about their role in disease.

"I'm really looking for people who think originally and are very open," says Carninci. "It's a new field, and we know almost nothing. So it's important to find people who always question the dogma of the day." ■

**Amy Maxmen** is a freelance writer in New York.

# COLUMN

# Time to reflect

A lab retreat provides a chance to rethink and advance the research programme, says **Eleftherios Diamandis**.

Labs sometimes struggle to stay competitive and energized, and periods of successful discovery may be followed by stretches filled with little more than attempts to address knowledge gaps or promote translation. These are essential processes, but innovation may suffer if such periods last too long.

As director of a 25-person research lab (ten graduate students, five technologists, six postdocs, a research coordinator and affiliates), I know that focusing too much on the day-to-day business of e-mails, manuscripts and grant applications can delay the introduction of new techniques and ideas. One way to inject enthusiasm and re-excite the team is through a lab retreat to discuss everyone's projects and work.

The lab director generally sets the agenda for a retreat, but he or she should consider involving other lab members. Getting everyone on board is important: lab staff should understand that this is not an exercise devoted to identifying the 'good' and 'bad' projects or people, but is rather an open conversation about long-term planning to determine who needs help and how the director and other lab members might give it to them. Retreats can help to identify new strategies and areas where lab members are duplicating efforts on the same questions.

The director should circulate instructions ahead of time, explaining what he or she expects to be covered in the presentations and discussion. Constructive criticism should dominate. Participants should understand that this is not merely an update on research progress, but is instead a soul-searching exercise. Lab members should be asked to consider self-assessment questions such as 'Am I innovating or imitating?', 'Will my results lead to significant publications if successful?', 'Are there other techniques I should be using?' and 'Am I being too risk averse?'. Each person should address the novelty of their project as opposed to simply 'what they're working on'.

The lab supervisor might open the retreat with some remarks on each of the lab's research focuses, addressing the self-assessment questions as they pertain to the entire lab. Each member can then give a short presentation (perhaps five to seven slides) on his or her project, with special emphasis on innovations and anticipated research impacts.

Questions and criticisms must go beyond those at a typical lab meeting, where most questions focus on specific experiments and



DRAWN IDEAS/ICON IMAGES/GETTY

technical details. At the retreat, questions should examine appropriate approaches and even whether a project is worth pursuing at all. This may reveal that hypotheses or strategies have weaknesses that need correction — or that suggest that a project should be abandoned.

To promote debate, the director should consider assigning one or two lab members the role of devil's advocate. At my lab's retreat last year, one postdoc presented his findings on the anti-cancer properties of cardiac glycosides — drugs used to treat heart failure and arrhythmia — in a model of pancreatic cancer. The opponents challenged him, noting that his effective drug concentrations were ten times higher than the safety limit in humans — it might be possible to treat the cancer, said one, but the patient would die of cardiac arrest. Discussion followed on ways of retaining the anticancer properties but avoiding the toxicity. Another critic suggested that it could be possible to find a drug that acts with the cardiac glycoside, enhancing its anticancer activity at safer doses. After the retreat, this idea led the postdoc to perform a high-throughput screen of a 10,000-molecule chemical library to find such an agent. He is now testing a drug combination in animal models.

Lab directors are responsible for proposing and contemplating the direction that a lab will take in future. But retreats help to foster an annual re-examination of projects. Members need to recognize that the success of a lab depends on their capacity to innovate. In the end, all will share the rewards. ■

**Eleftherios Diamandis** is professor of laboratory medicine and pathobiology at the University of Toronto in Canada.

# FOR YOUR INFORMATION

*A little knowledge ...*

BY CONOR POWERS-SMITH

**W**hen Allison came in, Jennifer muted the TV, and asked, "Well?"

Allison shrugged. Only when she was seated on the sofa, under Jennifer's smiling scrutiny, did Allison's determinedly fixed expression give way to a wide grin.

Jennifer's smile widened in sympathy with her roommate's. She prompted, "Good?"

"Yeah."

"What'd you do?"

"A movie. Walked around for a while." They could've been picking up trash by the side of the road for all Allison cared. She heard herself blurt, "I really like him."

"I seem to get that impression. But this was only date two, right?"

Allison felt like she'd known Kevin for years, her whole life, longer. Still, she had to nod. "But we're going out again tomorrow."

"So this is a thing now," Jennifer said. "Officially."

"I guess. I hope."

"So, don't you think it's time we checked him out?"

"Jen..."

"He's got a profile, hasn't he?"

"Yeah. You pretty much have to now, or you're some kind of weirdo."

"So...?"

"I don't know."

"Allie, it's not being nosy. It's right there for everyone to see. It's just being responsible. I mean, if this is going to be a thing, you owe it to yourself to go in with your eyes open."

"I guess. But —"

"And admit it: you're just the tiniest bit curious."

Allison could hardly deny that. "Well... all right."

Jennifer sat in front of the computer, Allison behind her, on the edge of Jennifer's bed, craning her neck to see the screen. *Geno-Me* was slick and inviting, as befitting one of the world's most popular social media sites. The stylized double helices bordering each page glowed their soft greens and blues.

"Okay," Jennifer said. "The physical stuff's locked out, unless he's friended you." She glanced over her shoulder.

Allison shook her head. "We didn't talk much about it." In her mind, she was reviewing the few dismissive words they'd spoken on the subject, trying to determine whether they'd established a tacit understanding that neither would look at the other's profile.

Kevin had called it 'a marginally more accurate horoscope' but didn't people sometimes read horoscopes just for fun?

"The personality stuff's more fun anyway," Jennifer said. "Okay, let's see: H11Bβ, that's cool. He has a good sense of humour?"

"Definitely."

"But he has his serious side," Jennifer said. Her hand and wrist were busy working the mouse. As the cursor slid across the 23 pairs of rather cartoonish chromosomes on the screen, small sections lit up, boxes of text appeared beside them, and other sections, seemingly unrelated, glowed in sympathy. "And — Uh oh."

"What?"

"HOPPER9, B2F11, WELLER-WYMAN and no 17J-CROSSHAIRS. That's got stubbornness written all over it. Have you noticed that?"

"No. I mean, he didn't want to see the movie I did. But I didn't care."

"Well, you better get used to that. He'll be picking the movies."

Allison couldn't imagine a thing like that bothering her. Still, it was disconcerting to find a stain, however small, on her previously spotless conception of Kevin.

"And, look, Allie: NICEL7. That'd be okay if he had JIB4, or the elongated DONALDSON-HARVEY, but, nope. That means highly confrontational. *Highly*."

"Huh," Allison said, remembering the business of Kevin shushing the pair of obnoxious teenagers who'd been talking during the movie. But wasn't it admirable to stand up for one's self, when the situation dictated? Then again, the kids hadn't been so loud.

"Oh, Allie: 76UNION-Y-SAIL. Hostility towards authority. Look, that's linked to oppositional disorder ... all kinds of things ... *crime*."

Allison could come up with nothing to corroborate that, which somehow made it worse. She sounded unconvincing even to herself when she said, "But you can't tell how these things are going to express themselves. Environment's just as important. Right? Free will? What's the saying? It's a list of ingredients, not a recipe."

"Yeah, but look at the ingredients. You can't make a cake with ... sawdust and broken glass."



When Allison came in, Jennifer set aside her magazine, and said, "So?"

Allison's stoical expression faltered before she reached the sofa. By the time she was seated beside Jennifer, her mouth was bent in a steep frown, and her eyes were welling up.

"It's over."

"Oh, Allie. You broke it off? It's probably for the best. Considering."

"He did."

"What? Don't tell me he didn't like *your* profile. That's some nerve, with —"

"He never looked at my stupid profile. But he knew I looked at his. He said he could tell right away, I was treating him differently. I could hear myself doing it, too, but I couldn't stop. It's like, when you know those things about someone ... when you *think* you know ..."

"What was it, some kind of trap? Remember, he has VIKING-F11? They list that as strategic thinking, but one of the corollaries is deceit."

"It's a trap, but he didn't set it. The whole thing's a trap."

"Was he very ... confrontational?"

"Stop it, Jen. It wasn't even him we were looking at."

"We got the wrong profile?"

"It was his profile, but not his genome. He posted a different one. One of those historical reconstructions they do."

"What? Who?"

"Gandhi."

Jennifer stared. Finally she said, "Well, I'd call that deceitful. I bet he has VIKING, at least."

Allison stared back for a moment, then stood. "I'm going to bed." ■

**Conor Powers-Smith** lives on Cape Cod in Massachusetts, where he works as a reporter and writes fiction in his spare time.